ERRORS OF MEASUREMENT AND CORRELATION

BY
EDWARD E. CURETON, Ph.D.

ARCHIVES OF PSYCHOLOGY
R. S. WOODWORTH, Editor

No. 125

NEW YORK
May, 1931

ERRORS OF MEASUREMENT AND CORRELATION

BY

EDWARD E. CURETON

ARCHIVES OF PSYCHOLOGY
R. S. WOODWORTH
Editor

No. 125

NEW YORK
May, 1931

ACKNOWLEDGMENTS

I am indebted chiefly to Dr. Edward L. Thorndike, who has sponsored this study, for advice, encouragement, and inspiration. Dr. Henry E. Garrett, Dr. Henry A. Ruger, and Dr. Helen M. Walker have contributed invaluable advice during the progress of the investigation. I am indebted to Mr. Jack W. Dunlap, Dr. Laura H. V. Kennon, Dr. Irving Lorge, and Dr. Ella Woodyard for reading the manuscript and offering many helpful suggestions.

My interest in the problems discussed is due mainly to Dr. Truman L. Kelley of Harvard University. The pioneer work of Professor C. Spearman of the University of London has laid the foundations for all subsequent investigations in correlational psychology.

EDWARD E. CURETON
April 10, 1931

CONTENTS

APPENDIX

CHAPTER I

The Problem

Educators and psychologists often employ statistical formulas in the solution of their problems without noting the assumptions on which these formulas are based. This situation is particularly true with regard to the application of the methods of correlational analysis to data obtained from mental and educational tests. The theory of correlation, as developed principally by the work of the English Biometric School, was intended to apply to measurable physical characters. In order to apply it to test scores, the investigator must satisfy himself that these scores possess certain of the attributes of such measurements. Assumptions so obvious as to need no explicit statement when applied to measurements, such as equality of units at different parts of a scale, and independence of observational errors and values of the variable measured, require careful examination and verification when applied to test scores.

The first problem of the present investigation is to examine into the assumptions which underlie the principal formulas of correlational psychology. Some of these assumptions will be found to be fundamental necessary conditions for the applicability of the formulas. In such cases it is necessary to point out the limitations in methods of test construction and application implied, and this is the second problem. Some, on the other hand, which have been employed extensively in the past, will be found to be unnecessary under certain experimental conditions. The third problem, then, is to discover these conditions wherever possible, and to derive modified formulas applicable under them.

The consistency of data with assumptions can often be checked by statistical tests applied to the data. The devising of such checks is a fourth problem. Finally, if formulas are to be of any considerable usefulness, their standard errors should be known. The derivation of these standard errors for the principal formulas of correlational psychology is the fifth and last problem to be dealt with in this study.

CHAPTER II

RELIABILITY

Notation. A number of mental traits are to be measured by fallible tests. Each test will be supposed to consist of two, three, or more forms, A, B, C, etc. The values of the underlying traits, considered as though measured without error, will be designated X_∞, X_ω, X_γ, X_η, etc. The subscripts of the corresponding scores on Form A will be Arabic numerals, on Form B, small Roman numerals, and on Form C, large Roman numerals. The letter M will designate a mean, x a score taken as a deviation from the mean, so that $x = X - M$, and σ a standard deviation. We then have the following system of notation:

True Value of Trait	Form A score	Form B score	Form C score	Form A mean	Form A standard dev.
X_∞	X_1	X_i	X_I	M_1	σ_1
X_ω	X_2	X_{ii}	X_{II}	M_2	σ_2
X_γ	X_3	X_{iii}	X_{III}	M_3	σ_3
X_η	X_4	X_{iv}	X_{IV}	M_4	σ_4

etc.

This system of notation uses only a single subscript to designate any of the quantities involved, and it may readily be extended to any number of variables.

Sources of error in mental tests. When tests are used in an attempt to secure an estimate of the magnitude of some mental trait in an individual, there are at least four sources of error. First, the test items may sample systematically some trait other than the one they are designed to measure. Second, the test items may fail systematically to sample some important aspect of the trait they are designed to measure. These first two sources of error are the essential problems of validity, and they need not concern us at present. We shall assume that *the mental trait measured by the several forms of a test is whatever mental ability or combination of abilities causes simultaneous variation in the scores on these forms.**

* A discussion of this and other assumptions regarding the fundamental definition of reliability is given by Kelley (1924).

The third source of error resides in the fact that the given test situation is only a single sample of all possible situations calling for the exercise of the test ability. The individual changes from hour to hour and from day to day in a manner that may affect his test performance without changing his underlying ability. It might at first seem possible to estimate the magnitude of this individual variability—or rather its average magnitude in a group—by applying the same test at different times and noting the discrepancies between successive scores. The only reasonable way to do this, in order to eliminate the disturbing effects of memory of the previous responses to specific test items, would be to wait until these responses had been forgotten, and the time for this would normally be so great that it would no longer be safe to assume that there had been no significant change in the underlying trait during the interval.

The fourth source of error lies in the fact that any single test includes only a limited sample of the possible number of items measuring the trait under consideration. The magnitude of this error may be estimated by comparing the scores on random halves of the test, i.e., by giving the items of Forms A and B simultaneously. Errors of this fourth type will be called test errors hereafter, and errors of the third type, response errors. The two sets of errors taken together constitute the errors of measurement, when we are considering only the reliability, and not the validity of the test.

Fundamental factor pattern. A factor pattern is an equation or set of equations expressing the definitions and assumptions regarding the make-up of test scores in terms of abilities and errors in any given hypothetical situation. For the case of tests measuring the same fundamental ability, we may think of this ability as unitary, whence we have,

$$x_1 = f(X_\infty, \Delta_1),$$

where Δ_1 is the error of measurement in X_1. If X_1 is related to X_∞ and Δ_1 in an approximately linear manner, or if the variation in the values of X_1 over the group in which it is measured is small in comparison to the values themselves, no matter what the form of the function, we may write, to a close approximation, transferring origins to the means,

$$x_1 = c_1 x_\infty + \delta_1.$$

The value c_1 is a constant designating the ratio of the units in which x_1 is measured to the units in which x_∞ is measured. We may assume without loss of generality that δ_1 is measured in the units of x_1. Then, considering two forms of the test, we may write,

$$x_1 = c_1 x_\infty + \delta_1, \qquad x_i = c_i x_\infty + \delta_i. \tag{1}$$

Fundamental factor pattern for two tests that measure the same ability.

The factors c_1 and c_i are constants descriptive of the units in which x_∞ is measured in the two forms of the test, and δ_1 and δ_i are the respective errors of measurement. If we assume that these errors of measurement are uncorrelated with each other and with the true abilities of the subjects in any given group, as is implied in our definition of the trait measured, we may write,

$$\sigma_1^2 = c_1^2 \sigma_\infty^2 + \sigma_{\delta_1}^2, \qquad \sigma_i^2 = c_i^2 \sigma_\infty^2 + \sigma_{\delta_i}^2, \tag{2}$$

and,

$$\Sigma x_1 x_i / N = \Sigma (c_1 x_\infty + \delta_1)(c_i x_\infty + \delta_i)/N = c_1 c_i \sigma_\infty^2. \tag{3}$$

Definition of reliability. Since $c_1^2 \sigma_\infty^2$ is less than σ_1^2 by an amount equal to $\sigma_{\delta_1}^2$, we may define the reliability of the test as the ratio of the true variance (squared standard deviation) to the obtained variance. This definition, it should be noted, is quite general. It implies only the previous definition of the mental trait measured. It does not assume that the errors of measurement will sum to zero, nor that the several forms of the test are equally variable or reliable. As a pure definition, it does not in fact involve any other form of the test than the one under consideration. It assumes only that in the group taken as the sample, there is no correlation between the magnitude of the trait and that of the error of measurement. Calling the reliability coefficient so defined R_1, we have,

$$R_1 = c_1^2 \sigma_\infty^2 / \sigma_1^2. \tag{4}$$

Definition of the reliability coefficient of one form of a test.

Statistical estimation of reliability. The value of σ_1^2 may be obtained directly from the data. The value of $c_1^2 \sigma_\infty^2$ is unknown, but it may be estimated by introducing certain additional assump-

tions. If we let $\Sigma x_1 x_i / N = p_{1i}$ (covariance of x_1 and x_i*), we obtain at once from (3),

$$c_1 c_i \sigma_\infty^2 = p_{1i}. \tag{5}$$

To find the value of $c_1^2 \sigma_\infty^2$, we must evidently multiply this by the ratio c_1/c_i.

It is ordinarily impossible to determine with accuracy how the units of measurement of one form of a test compare with those of another form. In a fairly large sample we might assume that an error of measurement is equally likely to be positive or negative, so that the sum of all such errors in the sample would approach zero. Then, since,

$$M_1 = c_1 M_\infty + M_{\Delta_1}, \text{ and } M_i = c_i M_\infty + M_{\Delta_i},$$

$c_1/c_i = M_1/M_i +$ (terms of the order of M_{Δ_1}/M_i and terms of higher orders, all of which are negligible in comparison with M_1/M_i if M_{Δ_1} and M_{Δ_i} are close to zero).

This assumption, however, is quite dangerous. It is not unlikely that either form of the test will contain unique non-chance elements, which in spite of being non-vanishing in the group are properly to be classed with the errors of measurement because of their irrelevance. This point has been more fully discussed by Kelley (1924). Furthermore, the values of M_1 and M_i in ratio comparisons of this sort must be measured from true zero-points of the underlying abilities, a condition at best only approximated by a very few mental and educational tests. And this very approximation is based on a further assumption. In the scaling of a test, the standard deviation is ordinarily taken as the unit of measurement. But it is obvious from (2) that the magnitude of the standard deviation depends in part upon the size of the errors of measurement. Hence, a mean, expressed in terms of some multiple of the corresponding standard deviation as a unit, cannot be compared with another mean so expressed, except on the assumption that the errors of measurement are proportional to the units, i.e., that the fundamental reliabilities of the two forms of the test are equal. For example, suppose that the basic units of measurement are equal in the two forms. Then in (2),

* The covariance is the first product-moment coefficient of two sets of observations.

$c_1^2\sigma_\infty^2 = c_i^2\sigma_\infty^2$, but σ_1^2 will not equal σ_i^2 unless $\sigma_{\delta_1}^2 = \sigma_{\delta_i}^2$ also. The form of the test having the greater error of measurement will have the greater standard deviation, and as a consequence its mean, measured in standard units, will be lower than that of the more reliable form. An important corollary of this fact may be stated as follows: *Standard scores are comparable measures only in case the measurements so compared are of equal reliability.*

There is one method of estimating the ratio c_1/c_i, however, that is based entirely on averages. This method assumes that the difference between the means of two groups will be the same whichever of two fallible tests is used to determine this difference. These means might be successive age or grade averages taken from the test norms. If we choose some age or grade range such that from 1/6 to 1/20 of the experimental group falls outside it at either end, the mean score increments on the two forms of the test corresponding to the given age or grade increment will be approximately equivalent, and their ratio will be equal to c_1/c_i.

If our sample is large, we may obtain the ratio c_1/c_i without recourse to norms based on other groups, and so avoid the assumption, implicit in the procedure of the previous paragraph, that norms are available for the two forms of the test, obtained from comparable groups. Giving each individual in the experimental group a total score equal to the sum of his scores on the two forms, we may secure the sub-groups by taking the lowest and highest quarters of the original groups*. The ratio of the differences between the means of the sub-groups on the two tests may then be taken as equal to c_1/c_i for the given total group. When we know this ratio, we may substitute at once in (4), and we obtain,

$$R_1 = c_1^2\sigma_\infty^2/\sigma_1^2 = (c_1 c_i \sigma_\infty^2/\sigma_1^2)(c_1/c_i),$$

and from (5),

$$R_1 = p_{1i}c_1/\sigma_1^2 c_i = b_{i1}c_1/c_i. \qquad (6)$$

Reliability of Form A.

Similarly,

$$R_i = p_{1i}c_i/\sigma_i^2 c_1 = b_{1i}c_i/c_1. \qquad (7)$$

Reliability of Form B.

* Strictly speaking, we should use the highest and lowest 27 per cent, as Kelley has proved that in this case the ratio of the difference between the means to its standard error will be a maximum, the distribution of scores in the total group being normal. See Jensen (1928), p. 361.

If $c_1 = c_i$, these become simply,

$$R_1 = b_{i1} \tag{8}$$
$$R_i = b_{1i} \tag{9}$$

Reliabilities of the two forms of a test when the units of measurement are equal.

If $c_1/\sigma_1 = c_i/\sigma_i$, we obtain from (4) and (5),

$$R_1 = R_i = r_{1i} \tag{10}$$

Reliabilities of equally reliable tests.

The value of r_{1i} is the ordinary reliability coefficient. This has often been supposed to apply properly only to comparable tests —tests measuring the same ability in the same units with the same error, and whose standard deviations are therefore equal. It is seen here to have much wider usefulness than this, being applicable wherever we have two equally reliable tests of the same ability, no matter what the units of measurement may be in either case. This fact was pointed out by Kelley (1924).

Interpretation of the reliability coefficient. It is of interest to note that the reliability coefficient is not only the ratio of the variance of the true scores to that of the obtained scores, but also the square of the correlation between the obtained scores and the corresponding true scores, if this is linear. For,

$$\begin{aligned} r_{1\infty} &= \Sigma x_\infty (c_1 x_\infty + \delta_1)/N\sigma_\infty\sigma_1 \\ &= c_1\sigma_\infty^2/\sigma_\infty\sigma_1 = c_1\sigma_\infty/\sigma_1. \end{aligned}$$

But

$$R_1 = c_1^2\sigma_\infty^2/\sigma_1^2,$$

So that

$$R_1 = r_{1\infty}^2. \tag{11}$$

And

$$r_{1\infty} = R_1^{1/2}. \tag{12}$$

This last expression has been termed the index of reliability*.

Reliability of the sum of the scores on two forms of a test. If two forms of a test have been given, we may determine

* Monroe (1923), p. 206, gave this name to the square root of the reliability coefficient, ascribing it to Kelley, who had actually used it in a paper, but with no intention of coining a term. Monroe appears to have been the first to do this. See Walker (1929), p. 117, for further discussion of this point.

the reliability of the total score obtained by adding together the scores on the separate forms. For since,

$$x_{(1+i)} = x_1 + x_i = (c_1 + c_i)x_\infty + \delta_1 + \delta_i,$$

$$R_{(1+i)} = (c_1 + c_i)^2 \sigma_\infty^2 / \sigma_{(1+i)}^2$$

$$= (c_1^2\sigma_\infty^2 + c_i^2\sigma_\infty^2 + 2c_1c_i\sigma_\infty^2)/(\sigma_1^2 + \sigma_i^2 + 2p_{1i}).$$

$$R_{(1+i)} = [p_{1i}(c_1/c_i + c_i/c_1 + 2)]/[\sigma_1^2 + \sigma_i^2 + 2p_{1i}]. \tag{13}$$

Reliability of Form A plus Form B.

If $c_1 = c_i$,

$$R_{(1+i)} = 4p_{1i}/(\sigma_1^2 + \sigma_i^2 + 2p_{1i}). \tag{14}$$

Reliability of Form A plus Form B when the units of measurement are equal.

If in addition $\sigma_1^2 = \sigma_i^2$, we obtain, on dividing numerator and denominator by $\sigma_1\sigma_i$,

$$R_{(1+i)} = 2r_{1i}/(1 + r_{1i}). \tag{15}$$

Reliability of Form A plus Form B when these are comparable forms.

This last equation is the well-known Spearman-Brown formula for the reliability of a test twice as long as either of the forms used in computing the original reliability coefficient. It applies strictly only to comparable tests, but Kelley has shown (1924) that whenever the tests are of approximately equal reliability, and the units of one are not very much greater than those of the other (more than twice as great, say), it still applies to a very close approximation.

Reliability more accurately determined from three forms. If three forms of a test have been given to the same group, we may determine the reliability of any one of them without recourse to the somewhat dubious methods of evaluating the ratio c_1/c_i described above. For, as in (1), (2), (3), and (5),

$$x_1 = c_1x_\infty + \delta_1. \quad \sigma_1^2 = c_1^2\sigma_\infty^2 + \sigma_{\delta_1}^2. \quad p_{1i} = c_1c_i\sigma_\infty^2.$$

$$x_i = c_ix_\infty + \delta_i. \quad \sigma_i^2 = c_i^2\sigma_\infty^2 + \sigma_{\delta_i}^2. \quad p_{1I} = c_1c_I\sigma_\infty^2.$$

$$x_I = c_Ix_\infty + \delta_I. \quad \sigma_I^2 = c_I^2\sigma_\infty^2 + \sigma_{\delta_I}^2. \quad p_{iI} = c_ic_I\sigma_\infty^2.$$

From the three right-hand expressions,

$$p_{1i}p_{1I}/p_{iI} = c_1^2\sigma_\infty^2, \tag{16}$$

and from (4),

$$R_1 = c_1^2\sigma_\infty^2/\sigma_1^2 = p_{1i}p_{1I}/p_{iI}\sigma_1^2,$$

or,

$$R_1 = r_{1i}r_{1I}/r_{iI}. \tag{17}$$

Reliability of Form A from a knowledge of three tests of the same ability.

Similarly,

$$R_i = r_{1i}r_{iI}/r_{1I}, \tag{18}$$

and,

$$R_I = r_{1I}r_{iI}/r_{1i}. \tag{19}$$

The square roots of quantities such as those given in (17), (18), and (19) have been given by Spearman (1927), Appendix p. xvi, as the correlations between the respective tests and g, the general factor common to them, when the theory of two factors holds. These equations express this relationship for the special case in which the specific factors can be taken entirely as errors of measurement.

For the reliability of the total score obtained by adding together the scores on the three forms, we have,

$$R_{(1+i+I)} = (c_1 + c_i + c_I)\sigma_\infty^2/\sigma_{(1+i+I)}^2$$

$$\begin{aligned} R_{(1+i+I)} = (&p_{1i}p_{1I}/p_{iI} + p_{1i}p_{iI}/p_{1I} + p_{1I}p_{iI}/p_{1i} \\ &+ 2p_{1i} + 2p_{1I} + 2p_{iI})/(\sigma_1^2 + \sigma_i^2 + \sigma_I^2 \\ &+ 2p_{1i} + 2p_{1I} + 2p_{iI}). \end{aligned} \tag{20}$$

Reliability of Form A plus Form B plus Form C.

Reliability of the sum of several tests. Formula (20) may be generalized to give the reliability of the sum of any number of tests of the same function. We shall have to change our notation in dealing with more than three tests. Calling the forms X_1, X_2, . . .X_n, and the reliability coefficient R_n instead of $R_{(1+2+\ldots+n)}$, we have for n forms,

$$R_n = (c_1 + c_2 + \ldots + c_n)^2 \ \sigma_\infty^2/\sigma_{(1+2+\ldots+n)}^2.$$

In estimating the value of $c_1^2\sigma_\infty^2$, we may take as equally valid the values $p_{12}p_{13}/p_{23}$, $p_{12}p_{14}/p_{24}$, . . .$p_{1(n-1)}p_{1n}/p_{(n-1)n}$, and the final

estimate will be the average of all the values so obtained, $(n-1)(n-2)/2$ in number. But we must also determine $c_2^2\sigma_\infty^2$, $c_3^2\sigma_\infty^2$, . . . $c_n^2\sigma_\infty^2$ by the same process. It is evident, therefore, that the numerator will contain all the possible values of $p_{jk}p_{jq}/p_{kq}$, where j, k, and q take all possible combinations of values from 1 to n except one another. We obtain therefore,

$$R_n = [2S'(p_{jk}p_{jq}/p_{kq})/\{(n-1)(n-2)\} + 2Sp_{jk}] \div [\Sigma\sigma_j^2 + 2Sp_{jk}]. \tag{21}$$

Reliability of the sum of n tests of the same function.

Where Σ is a summation from 1 to n,

S is a summation from 1 to $n(n-1)/2$, and

S' is a summation from 1 to $n(n-1)(n-2)/2$.

Criterion of equal units of measurement. If the several forms of the test are measured in the same fundamental units but differ in their reliabilities; i.e., if $c_1 = c_2 = \ldots = c_n$, but $\sigma_1, \sigma_2, \ldots \sigma_n$ are different, then,

$$c_1c_2\sigma_\infty^2 = c_1c_3\sigma_\infty^2 = \ldots = c_1c_n\sigma_\infty^2 = \ldots = c_{(n-1)}c_n\sigma_\infty^2,$$

whence,

$$p_{12} = p_{13} = \ldots = p_{1n} = \ldots = p_{(n-1)n}.$$

This may be stated, *If several tests of the same function measure that function in the same basic units, the covariances of these tests will all be equal, except by chance.* We may treat the forms of the test as successive samplings of the ability of the group, and compare the difference between any two covariances with the standard error of this difference. If there are a considerable number of forms, more than 6 or 7, say, we may compute the standard deviation of the actual distribution of covariances, and compare this with the theoretical standard error of a random covariance of the mean order of magnitude of those observed. Formulas for such comparisons will be given in Chapter VI. These comparisons permit the investigator to determine when it is necessary to use a full formula, such as (20) or (21), and when it is reasonable to use the simplified formulas next to be considered.

Special cases of reliability of the sum of several tests. From (21), assuming all covariances equal,

$$R_n = n^2p/(\Sigma\sigma_j^2 + n(n-1)p). \tag{22}$$

Reliability of the sum of n tests of the same ability, measured in the same units.

As before, Σ is a summation from 1 to n, and p is taken here as the mean covariance. If in addition $\sigma_1^2 = \sigma_2^2 = \ldots = \sigma_n^2$, within the standard errors of their differences,

$$R_n = np/(\sigma^2 + (n-1)p). \tag{23}$$

where σ^2 is the mean variance. Then on dividing numerator and denominator by σ^2,

$$R_n = nr/(1 + (n-1)r). \tag{24}$$

Reliability of the sum of n comparable tests.

This is the Spearman-Brown formula, used to estimate the reliability of the sum of a number of comparable tests, when all the tests have actually been given to a group. The value of r should be taken as the average intercorrelation. The computation of this value has been discussed by Edgerton and Toops (1928).

Estimated reliability of one form of a test from a knowledge of several. For this purpose we have as a generalization of (17),

$$R_1 = 2S''(r_{1j}r_{1k}/r_{jk})/(n-1)(n-2). \tag{25}$$

Reliability of Form A, from a knowledge of n tests of the same ability.

The symbol S" represents a summation from 1 to $(n-1)(n-2)/2$. If $p_{12} = p_{13} = \ldots = p_{1n} = \ldots = p_{(n-1)n}$, so that we may assume that $c_1 = c_2 = \ldots = c_n$, we have as a generalization of (8),

$$R_1 = \Sigma b_{j1}/(n-1) \tag{26}$$

Reliability of Form A from a knowledge of n tests of the same ability measured in the same units.

The case in which we have comparable tests will be discussed later.

Practical implications of assumptions. All of the formulas for reliability rest on the assumption that the errors of measurement are uncorrelated. The mean square error, and its complement, the reliability coefficient, are estimated from the ratio of

the true variance to the obtained variance. The latter will always contain both the response error and the test error. But the experimental conditions must be relied upon to insure that our estimate of the true variance is free from any *systematic* error of either sort.

Consider the response error. The several forms of the test must be given at such intervals that in general no important aspect of this error is necessarily present at successive testings. It is obvious that they should not be given at the same sitting. They should probably be given at least one, and perhaps several, days apart. But as the time between successive testings is lengthened, the response error merges gradually with progressive changes in the underlying ability itself. The time should be short enough to warrant the assumption that there has been no significant growth or learning during the interval, but long enough to insure that there are no important elements of the response error that persist from one testing to the next. It should also be of such duration as to avoid recurrent response errors. Thus the forms should probably not be given at the same hour of the day, nor perhaps even on the same day of the week. There is no mathematical criterion here, but the one who uses tests, if he desires to estimate their reliabilities, must keep in mind the nature of these errors in planning the appropriate intervals for any particular testing program.

The practice of giving a single test to a group, taking the odd and even questions (or any other combination) as two forms, computing their correlation, applying the Spearman-Brown formula, and assuming that the resulting coefficient is a valid estimate of the reliability of the test, is in error. Most if not all of the response error for each individual will be common to both "forms," and the resulting correlation between them will therefore be exaggerated. The difference between this coefficient and unity, in fact, might well be defended as a measure of the unreliability of the test due to test errors alone, and the difference between it and one obtained from the whole test and another comparable form given after a suitable interval might be taken as the unreliability due to response errors alone. Woodyard (1926), however, presents evidence that the response error is small as compared to other errors in tests. She states (p. 3),

"The most general conclusion from a review of the evidence is that the time factor is of small moment in causing an individual

to vary in the mental work he produces under conditions found in the administration of such standard tests as are common in intelligence and educational testing. In practically all of the data examined, whatever correlation is obtained for the short time interval is changed by but a few points in the second decimal place when the time interval is increased."

When the test material is of such a nature that examination with one form constitutes significant practice for the subject, tending to raise his scores on subsequent forms appreciably, we have a special problem. Any correlation between ability and improvability will introduce correlated response errors. This matter has been discussed by several writers, notably Spearman (1910), Brown (1910 exp.) and (1913), and Wilton (1914). The last-named writer finds that these errors can be eliminated by using an odd number of forms—at least five are necessary—given at successive equal intervals. He takes the sum of the scores on the even forms as Form A, and the sum of the scores on the odd forms as Form B, multiplying the first and last of these by 0.5.

The "halo" error in scoring subjective tests, comparing specimens of work with quality scales, and making trait ratings, may be treated as a part of the response error in estimating reliability. The reliability of an essay test can be determined by giving some of the questions on one day and some on another, and having a different person mark each set. The two sets then become the two forms of the test. In the case of the quality scale, the subject should submit two specimens of his work produced on different occasions. These should then be compared with the scale by different people. Trait ratings should be obtained from raters whose contacts with the person rated have been largely at different times, if this is possible. In the last two cases, there will still remain the scale error—the error resulting from the fact that both raters use the same specific set of scaled specimens or rating devices. No author to date has apparently thought it necessary to provide duplicate forms for a quality scale or rating scale, and no one yet knows to what extent two scales can differ in external appearance and still remain equally valid. Finally, there is the error involved in assuming that the raters are equally competent. The reliability of a rating depends upon the reliability of the rater as well as upon those of the person rated and the scale used.

The test error is a matter that concerns chiefly the author of the test. If it is to be uncorrelated with the true ability, the test

must be equally reliable throughout its useful range, that is to say, the test errors of those who make the higher scores must be neither greater nor less, on the whole, than the errors of those who make the lower scores. This implies that if the questions be arranged in order from easy to hard, the increments of difficulty must in general be equal, or at least that these increments must not exhibit any progressive change. For if, say, the increments of difficulty are greater in the upper ranges of the test than in the lower, the test errors will for that reason be higher for those who make high scores than for those who make lower scores. Furthermore, the intrinsic excellence of the easy questions must in general be equal to that of the harder ones, for the same reason. It is conceivable that a test might be constructed in such a manner that a systematic change in the intrinsic excellence of questions, going from easy to hard, might just be balanced by a systematic change in the increments of difficulty, proceeding in the opposite direction, but such a balance is not likely to be achieved in practice.

From the standpoint of reliability, the intrinsic excellence of a question may be judged roughly by its biserial correlation with the total test score. Strictly speaking, this intrinsic excellence should be defined as the partial biserial regression of the particular question on the total score, eliminating all the other questions, but the computation of such a regression coefficient involves the fourfold point-correlation of every question with every other question, a task which is in practice prohibitive. The specific excellence of the question—its biserial correlation with an outside criterion—may be taken in most cases as equivalent to its intrinsic excellence for scaling purposes.

A speed test should consist of questions of approximately equal difficulty, or if this is impossible, the easy and hard questions should be interspersed in random order rather than arranged in order of difficulty, since hard questions represent greater increments of difficulty than do easy ones from the standpoint of speed, if the test is scored according to the number correctly answered. In fact, it would appear that the only tests in which the questions should properly be arranged in order from easy to hard are pure power tests—perhaps only those designed to be applied without time limits.

Comparable tests. Let us assume that we have only two forms of a test; that these forms consist of questions of approx-

imately equal intrinsic excellence; that on each form there are about the same number of easy and hard questions, so that the increments of difficulty may be assumed to take no systematic trend, or (if the test has a time limit short enough to cause noticeable differences between scores achieved under this limit and scores achieved when there is no such limit) that the questions do not vary greatly in difficulty and are arranged in random order in this respect; and that the number of questions on each form is fairly large, say 50 or more. In this case any observed discrepancy between the mean score differences of the two forms, obtained from groups or sub-groups of different ability, will probably give no useful information regarding the relative magnitudes of the units of measurement in the forms, and we may just as well take these units as equal or as proportional to the corresponding standard deviations. Small differences in mean score increments between successive ages or grades are probably in this case to be attributed more to differences between the groups used in deriving the norms than to differences in the units of measurement, and slight discrepancies in the mean score differences between the upper and lower quarters of a group may be attributed to errors of measurement, unless the group is quite large. Hence for practical purposes, formulas (10) and (15) may be recommended for estimating the reliabilities of fairly comparable tests. If three or more forms are available, formulas (17), (18), (19), (20), (21), and (25) are still to be preferred, even though the forms are fairly comparable. Cases in which tests measure in the same basic units but with markedly different reliabilities are likely to be rare in practice. For this reason formulas (8), (9), (14), (22), and (26) are probably of theoretical interest only.

A reasonable check on the comparability of tests is to compare their variances and covariances, which should all in this case be equal respectively to one another. This is not a rigorous criterion, since it is possible, though highly improbable, that the forms might differ in their fundamental units of measurement in one direction and in their test errors in the other in such a manner as to make their variances equal. If all the covariances are equal, the tests necessarily measure in the same basic units. Tests may be comparable even though their means are different, provided that there are no scores close to zero or perfection on any form, and the variances and covariances are equal respectively to one another. For example, it would be quite simple to increase the

obtained mean score on one form by adding a number of questions so easy that practically everyone taking the test could answer them all correctly, without affecting its comparability to the other forms.

Estimated reliabilities of comparable tests. It is often necessary to predict the reliability of the sum of several comparable tests from a knowledge of a few, or to estimate the reliability of one form, without specifying which one, from a knowledge of several. From (23),

$$R_n = np_{1i}/((\sigma_1^2+\sigma_i^2)/2+(n-1)p_{1i}). \quad (27)$$

Reliability of the sum of n comparable tests estimated from a knowledge of two.

Assuming that $\sigma_1^2=\sigma_i^2$, and dividing numerator and denominator by this value,

$$R_n = nr_{1i}/(1+(n-1)r_{1i}). \quad (28)$$

This is the Spearman-Brown formula used as an instrument of prediction. The only difference between (28) and (24) is that the average intercorrelation in the latter is replaced by the single known correlation. Since predictions of this sort are possible only with comparable tests, (27) is only slightly superior to (28). Neither has any very definite meaning if σ_1^2 differs very markedly from σ_i^2.

Generalizing (27),

$$R_n = [2nSp_{jk}] \div [(m-1)\Sigma\sigma_j^2+2(n-1)Sp_{jk}]. \quad (29)$$

Reliability of the sum of n comparable tests, estimated from a knowledge of m of them.

The symbol Σ here represents a summation from 1 to m, and S a summation from 1 to $m(m-1)/2$. Assuming all variances and covariances equal, and dividing numerator and denominator by the variance,

$$R_n = nr/(1+(n-1)r). \quad (30)$$

This is the Spearman-Brown formula again. The value of r is to be taken here as the average intercorrelation among the m known forms of the test. Otherwise (30) is identical with (24) and (28). If we already know the value of R_m, we may write,

$R_n = [nR_m/m] \div [1 + ((n/m) - 1)R_m].$

$$R_n = nR_m/(m + (n - m)R_m). \tag{31}$$

Reliability of the sum of n comparable tests, estimated from a knowledge of the reliability of the sum of m of them.

This result was first published by Spearman (1910). It is the most general form of what is usually called the Spearman-Brown formula, although in this form it was published only by Spearman.

If $n = 1$ in (29), we obtain,

$$R_j = Sp_{jk}/((m - 1)\Sigma\sigma_j^2) \tag{32}$$

Reliability of one unspecified form of a test, estimated from a knowledge of m comparable forms.

Setting $m = 2$ in (32),

$$R_j = 2p_{1i}/(\sigma_1^2 + \sigma_i^2) \tag{33}$$

Reliability of one unspecified form of a test, estimated from a knowledge of two comparable forms.

Formula (33) should give a slightly better estimate than (10), the ordinary reliability coefficient. It differs from r_{1i} only in that its denominator is the arithmetic mean of the two variances instead of their geometric mean. Since with comparable tests the variances are approximately equal, this difference will be negligible in most practical situations.

Reliabilities of strictly comparable tests. If we have several tests which are closely comparable, and whose means in addition are approximately equal, we shall call these strictly comparable tests, to distinguish them from other comparable tests whose means are not necessarily equal. With such tests we may obtain somewhat better estimates of the reliabilities of unspecified single forms and of sums of several forms than are given by formulas (10), (15), (24), (27), (28), (29), (30), (31), (32), and (33). In this case we may assume that it is immaterial which questions occur in which forms of the test, and in what order any individual takes the several forms. If we have N individuals and two forms of the test given to each of them, we may then calculate a single mean and a single variance for the 2N measures, according to the formulas,

$$M = (\Sigma X_1 + \Sigma X_i)/2N. \tag{34}$$

$$\sigma^2 = (\Sigma X_1^2 + \Sigma X_i^2)/2N - M^2. \tag{35}$$

We may then calculate the covariance by the formula,

$$p = \Sigma X_1 X_i / N - M^2. \tag{36}$$

Substituting in either (33) or the usual product-moment formula,

$$R_j = r_1 = p/\sigma^2 \tag{37}$$

Reliability of a single form of a test, estimated from two strictly comparable forms, the particular form being unspecified.

The symbol r_1 (with a single subscript) will be used to designate a coefficient obtained from (34), (35), (36), and (37). This coefficient is the intraclass correlation. It may be substituted for the ordinary product-moment intercorrelation in (15) and (28), the two-variable and n-variable cases of the Spearman-Brown formula used to predict the reliability of the sum of several comparable tests from a knowledge of two, if these two are strictly comparable. It has a slightly smaller sampling error than the corresponding intercorrelation, and is therefore to be preferred whenever the several forms of the test are strictly comparable. If n such forms have been given to each of the N individuals, we may obtain in a manner similar to (34), (35), and (36),

$$M = (\Sigma X_1 + \Sigma X_2 + \ldots + \Sigma X_n)/nN. \tag{38}$$

$$\sigma^2 = (\Sigma X_1^2 + \Sigma X_2^2 + \ldots + \Sigma X_n^2)/nN - M^2. \tag{39}$$

$$p = 2(\Sigma X_1 X_2 + \Sigma X_1 X_3 + \ldots + \Sigma X_1 X_n + \ldots + \Sigma X_{(n-1)} X_n) \div (nN(n-1)) - M^2. \tag{40}$$

We have then,

$$R_j = r_n = p/\sigma^2 \tag{41}$$

Reliability of one unspecified form of a test, estimated from a knowledge of n strictly comparable forms.

The quantity r_n is the generalized intraclass correlation. It may be substituted for the average intercorrelation in formulas (24), (30), and (31), the variations of the Spearman-Brown formula for cases in which the scores on more than two forms of the test are known. Its sampling error is smaller than that of the corresponding intercorrelation. If n becomes at all large, the computation of p from (40) becomes exceedingly laborious. Furthermore, it has been shown by Fisher (1928), Ch. 7, that (41) gives

a slightly biassed estimate of the true value of r_n unless N is quite large as compared with n. Therefore r_n should in practice be computed by another method. This computation is discussed further in Appendix I.

CHAPTER III

Validity

General meaning of validity. The validity of a test is, broadly speaking, the efficiency with which it measures the trait it was designed to measure. This involves two errors, as noted previously, in addition to the test error and the response error. First, the test may contain a specific non-chance factor not found in the trait, and second, the trait may contain a specific non-chance factor not found in the test, i.e., the test may sample either more or less than the trait, or both. But in estimating the validity of a test we must also consider the test error and the response error, both in the test and in the criterion. A test is *fundamentally valid* if there is neither sort of specific non-chance factor present, but its *practical validity* depends also on its reliability. Suppose, for example, that we desire to predict success in college. This might be defined by the faculty as the point-hour ratio, which is obtained by multiplying the number of hours of A achieved by the student by 4, of B by 3, of C by 2, of D by 1, and of F by 0; and dividing the sum of these points by the number of hours per week carried. The trait which the test is designed to measure is then the sum-total of all systematic factors which influence the point-hour ratio. The latter may of course be influenced, and quite largely, by chance factors, which from their very nature are unpredictable. The reliability of the criterion may be obtained from the point-hour ratios of successive semesters or quarters by the methods outlined in the previous chapter. The fact that some applicants are refused admittance and that others drop out along the way introduces complications in the practical situation that may for present purposes be neglected.

Fundamental factor pattern for validity.

Let x_1 and x_i be scores on two forms of a test,

x_2 and x_{ii}, equivalent criterion measurements,

x_ω, the underlying ability measured by the test,

x_∞, the ability underlying the criterion measurements, which is the ability we are trying to measure. Then,

$$x_1 = c_1 x_\omega + \delta_1,$$
$$x_i = c_i x_\omega + \delta_i,$$
$$x_2 = c_2 x_\infty + \delta_2,$$
$$x_{ii} = c_{ii} x_\infty + \delta_{ii}.$$

If x_ω is correlated with x_∞, we may consider this correlation to be the result of a factor common to the corresponding measurements.

Let a be a factor common to all four measurements,
b, a factor common to x_1 and x_i but outside x_2 and x_{ii},
d, a factor common to x_2 and x_{ii} but outside x_1 and x_i.

We then have the following factor pattern,

$$x_1 = c_1 a + c_1 b + \delta_1,$$
$$x_i = c_i a + c_i b + \delta_i,$$
$$x_2 = c_2 a + c_2 d + \delta_2 = c_2 x_\infty + \delta_2,$$
$$x_{ii} = c_{ii} a + c_{ii} d + \delta_{ii} = c_{ii} x_\infty + \delta_{ii}.$$

The value x_∞ is, as stated above, the sum of all the abilities which cause systematic variation in the criterion scores. Since we have assumed that the two forms of the test measure the same fundamental abilities, the a and b factors in x_1 and x_i may be multiplied by the same c's; and similarly for the a and d factors in x_2 and x_{ii}. It is assumed that x_1 and x_i contain the same relative proportions of a and b; an assumption implied in the previous one, that the tests and criterion measures sample the same respective fundamental abilities, i.e., that x_ω is the same in the two tests and x_∞ is the same in the two criterion measurements. Making the additional assumptions implied in the factor pattern, namely that a, b, d, and all the δ's are uncorrelated with one another, we have in succession,

$$\left.\begin{aligned}
p_{1i} &= c_1 c_i \sigma_a^2 + c_1 c_i \sigma_b^2,\\
p_{12} &= c_1 c_2 \sigma_a^2,\\
p_{1,ii} &= c_1 c_{ii} \sigma_a^2,\\
p_{i2} &= c_i c_2 \sigma_a^2,\\
p_{i,ii} &= c_i c_{ii} \sigma_a^2,\\
p_{2,ii} &= c_2 c_{ii} \sigma_a^2 + c_2 c_{ii} \sigma_d^2 = c_2 c_{ii} \sigma_\infty^2.
\end{aligned}\right\} \qquad (1)$$

Furthermore,

$$\left.\begin{aligned}
p_{12} &= \Sigma(x_1(c_2 a + c_2 d + \delta_2))/N = c_2 p_{1a},\\
p_{i,ii} &= \Sigma(x_1(c_{ii} a + c_{ii} d + \delta_{ii}))/N = c_{ii} p_{1a},\\
p_{i2} &= \Sigma(x_i(c_2 a + c_2 d + \delta_2))/N = c_2 p_{ia},\\
p_{i,ii} &= \Sigma(x_i(c_{ii} a + c_{ii} d + \delta_{ii}))/N = c_{ii} p_{ia}.
\end{aligned}\right\} \qquad (2)$$

These equations hold only under the above assumption that a, b, d, and all the δ's are uncorrelated with one another. It is very important that the two forms of the test be given at such an interval that there will not be any recurrent response errors, which would introduce a spurious b-factor. The two criterion measures must likewise be free from any spurious d-factors. In the case of point-hour ratios, for example, we should not take the sum of the first-semester ratios as one form and the sum of the second-semester ratios as the other, since many courses run for a full year, and the "halo effect" in the second-semester marks in such cases would enter as a spurious d-factor. A better method would be to take the sum of the first, fourth, fifth, and eighth-semester ratios as one form, and the sum of the second, third, sixth, and seventh-semester ratios as the other.

Statistical estimation of fundamental validity. If the fundamental validity of the test is perfect, there will be no b-factor and no d-factor, and we obtain at once from equations (1),

$$p_{1i}p_{2,ii} = p_{12}p_{i,ii} = p_{1,ii}p_{i2}. \tag{3}$$

Criterion of perfect fundamental validity.

This is the well-known tetrad relation, expressed in terms of the covariances. If the fundamental validity is zero, there will be no a-factor, and the lower the fundamental validity, the smaller will be the variance of the a-factor in comparison with those of the b-factor and the d-factor. All the p's except p_{1i} and $p_{2,ii}$ must, therefore, become smaller, relatively to these two, as σ_a^2 becomes smaller relatively to σ_b^2 and σ_d^2. Hence, as the coefficient of fundamental validity we may write,

$$V_\infty = (p_{12}p_{i,ii}/p_{1i}p_{2,ii})^{1/2} = (p_{1,ii}p_{i2}/p_{1i}p_{2,ii})^{1/2}.$$

Combining the two right-hand expressions,

$$V_\infty = (p_{12}p_{1,ii}p_{i2}p_{i,ii})^{1/4}/(p_{1i}p_{2,ii})^{1/2}. \tag{4}$$

Coefficient of fundamental validity.

This value is a special case of the correlation corrected for attenuation. The roots are taken in order that the coefficient may vary with the variances themselves, rather than with their squares. The two right-hand expressions in the equation immediately preceding (4) should give the same value, within their sampling errors, according to our analysis. This gives a partial

check on the assumption that the two forms of the test and the two criterion measures sample respectively the same abilities. This essential equality may be more simply expressed,

$$p_{12}p_{i,ii}/p_{1,ii}p_{i2} = 1. \tag{5}$$

Partial check for equivalence of tests and criterion measures. This is a special case of the tetrad ratio.

The coefficient of fundamental validity will vary from zero for no fundamental validity to unity for perfect fundamental validity. The correlation coefficients may be used instead of the covariances in (4) and (5) without changing either its value or the value of its standard error.

Practical validity. Even if the fundamental validity of a test is perfect, its practical validity may still be very low, due to its lack of reliability. Practical validity may be defined in general terms as the accuracy with which a test measures the *ability underlying* a specified criterion. It should, therefore, vary with both the fundamental validity and the reliability of the test, but not with the reliability of the criterion measures. We must assume that the latter are fundamentally valid, however, i.e., that all systematic causes of variation in them (a-factors and d-factors) shall be included in x_∞. The practical validity is to be distinguished from the predictive value of the test, which depends on the reliability of the criterion measures as well as on the fundamental validity and reliability of the test.

Estimation of practical validity. It has been shown previously that the reliability of a test is equal to the square of its correlation with the fundamental ability underlying it. From the general definition of practical validity, and by analogy with the reliability coefficient, we may define the coefficient of practical validity as the square of the correlation between the *test* score and the fundamental ability underlying the *criterion* scores. Then from the factor pattern,

$$r_{1\infty}^2 = r_{1(a+d)}^2 = p_{1a}^2/\sigma_1^2\sigma_\infty^2.$$

From (2),

$$p_{12}p_{1,ii} = c_2c_{ii}p_{1a}^2,$$

and from (1),

$$p_{2,ii} = c_2c_{ii}\sigma_\infty^2,$$

so that,

$$V_1 = r_{1\infty}^2 = p_{12}p_{1,ii}/p_{2,ii}\sigma_1^2 = r_{12}r_{1,ii}/r_{2,ii}. \tag{6}$$

Coefficient of practical validity of Form A.

It may be seen that to determine the practical validity of a test we require only the one form, but we also require two criterion estimates. The latter need not be comparable, but their errors of measurement must be uncorrelated. The coefficient of practical validity resembles the square of Spearman's correlation between a test and the general factor (1927), Appendix p. xvi. But x_∞ in our case contains not only the general factor (here a), but also the group factor in the two criterion measures (here d). This shows clearly the fallacy involved in interpreting the function in this fashion in systems in which the theory of two factors does not hold.

If there is no d-factor, the coefficient of practical validity may be shown to be equal to the ratio of the variance of the common factor to that of the total test scores. Setting $d = 0$ in (1), we find,

$$p_{12}p_{1,ii}/p_{2,ii} = c_1^2\sigma_a^2,$$

and from (6)

$$V_1 = c_1^2\sigma_a^2/\sigma_1^2. \tag{7}$$

In the formulas for practical validity, no use has been made of x_i, these formulas being concerned only with a single test and the two criterion measures. Hence we have at once,

$$V_i = p_{i2}p_{i,ii}/p_{2,ii}\sigma_i^2 = r_{i2}r_{i,ii}/r_{2,ii}. \tag{8}$$

Practical validity of Form B.

$$\begin{aligned} V_{(1+i)} &= p_{(1+i)2}p_{(1+i)ii}/p_{2,ii}\sigma_{(1+i)}^2 \\ &= r_{(1+i)2}r_{(1+i)ii}/r_{2,ii}. \end{aligned} \tag{9}$$

Practical validity of Form A plus Form B in terms of combined scores.

Alternatively,

$$\begin{aligned} r_{(1+i)\infty}^2 &= (\Sigma x_\infty(x_1+x_i))^2/\sigma_\infty^2\sigma_{(1+i)}^2 \\ &= (p_{1a}+p_{ia})^2/\sigma_\infty^2\sigma_{(1+i)}^2 \ . \end{aligned}$$

From (2),

$$\begin{aligned} p_{12}p_{1,ii} &= c_2c_{ii}p_{1a}^2, \\ p_{i2}p_{i,ii} &= c_2c_{ii}p_{ia}^2, \\ p_{12}p_{i,ii} &= p_{1,ii}p_{i2} = c_2c_{ii}p_{1a}p_{ia}, \end{aligned}$$

so that,

$$(p_{1a}+p_{ia})^2 = (p_{12}p_{1,ii}+p_{i2}p_{i,ii}+p_{12}p_{i,ii}+p_{1,ii}p_{i2})/c_2c_{ii}.$$

Also from (1),

$$p_{2,ii} = c_2c_{ii}\sigma_\infty^2,$$

so that,

$$V_{(1+i)} = r^2_{(1+i)\infty} = (p_{12}p_{1,ii}+p_{i2}p_{i,ii}+p_{12}p_{i,ii} +p_{1,ii}p_{i2})/(p_{2,ii}(\sigma_1^2+\sigma_i^2+2p_{1i})). \tag{10}$$

Practical validity of Form A plus Form B in terms of the separate scores.

Significance of measures of validity. The practical validity of a test is the most important feature of its real usefulness. Even though the fundamental validity may not be perfect, the practical validity, due to the high reliability of the test, may well be greater than the reliability of the criterion. If the fundamental validity is perfect, the practical validity should equal the reliability of the test. It is important to note that in validating a test, the criterion need not be very reliable, provided it is fundamentally valid. Thus in the case of college grades, if the point-hour ratio is taken by *fiat* as the definition of academic success, a test might well be constructed whose reliability would be so high as compared to the reliability of teacher's marks, that it would be a better measure of the systematic mental traits underlying academic success than the total academic record—the point-hour ratio for the entire period of attendance—and this in spite of a fundamental validity definitely (though not greatly) less than perfect.

If the object of giving a test is simply to *predict* academic success as measured by the point-hour ratio, say, then its predictive value will be given by the raw correlation between the test scores and the criterion measures. This correlation is of course dependent upon the reliabilities of both the test and the criterion. Its value, as will be shown in the next chapter, can only approach the geometric mean of the two reliability coefficients as a maximum.

CHAPTER IV

Correction for Attenuation

Fundamental factor pattern. The basic problem of this chapter is the estimation of the correlation between the true abilities underlying two sets of test scores. If we have two forms of the test measuring each trait, we may assume the following factor pattern to hold:

$$x_1 = c_1 x_\infty + \delta_1.$$
$$x_i = c_i x_\infty + \delta_i.$$
$$x_2 = c_2 x_\omega + \delta_2.$$
$$x_{ii} = c_{ii} x_\omega + \delta_{ii}.$$

The problem is then to determine the value of $r_{\infty\omega}$. If we assume that of the values x_∞, x_ω, δ_1, δ_i, δ_2, and δ_{ii}, x_∞ and x_ω alone are correlated, we obtain,

$$\left.\begin{aligned} p_{1i} &= c_1 c_i \sigma_\infty^2, \\ p_{12} &= c_1 c_2 p_{\infty\omega}, \\ p_{1,ii} &= c_1 c_{ii} p_{\infty\omega}, \\ p_{i2} &= c_i c_2 p_{\infty\omega}, \\ p_{i,ii} &= c_i c_{ii} p_{\infty\omega}, \\ p_{2,ii} &= c_2 c_{ii} \sigma_\omega^2. \end{aligned}\right\} \tag{1}$$

Experimental implications. By giving all four of the tests at different times, the response errors could all, theoretically, be rendered independent. But this is in general impractical. If the Form A tests be given at one time and the Form B tests at another, however, it will still be possible to get an unbiassed estimate of $r_{\infty\omega}$. In this case, the response errors in x_1 and x_2 will be correlated, as will those in x_i and x_{ii}, but we may take δ_1 and δ_{ii} to be uncorrelated, and the same for δ_i and δ_2.

Statistical estimation of the correlation between underlying abilities. From the above considerations and from (1), we find,

$$p_{1i} p_{2,ii} = c_1 c_i c_2 c_{ii} \sigma_\infty^2 \sigma_\omega^2,$$
$$p_{1,ii} p_{i2} = c_1 c_i c_2 c_{ii} p_{\infty\omega}^2,$$

and,

$$r_{\infty\omega} = p_{\infty\omega}/\sigma_\infty \sigma_\omega = (p_{1,ii} p_{i2}/p_{1i} p_{2,ii})^{1/2}. \tag{2}$$

Alternatively, dividing numerator and denominator by $\sigma_1\sigma_i\sigma_2\sigma_{ii}$,

$$r_{\infty\omega} = (r_{1,ii}r_{i2}/r_{1i}r_{2,ii})^{1/2}. \tag{3}$$

Correlation corrected for attenuation.

This formula has been noted by Yule (1927), pp. 213-14, as being unaffected by correlated errors in δ_1 and δ_2 or δ_i and δ_{ii}. His derivation of it, however, was based on a factor pattern without the c's, assuming both measures of each trait to be in the same units. The above demonstration shows that the formula is in fact independent of this assumption.

If we assume that all the errors are uncorrelated, as would be reasonable if all the tests had been given at different times, we obtain from (1),

$$p_{12}p_{i,ii} = c_1c_ic_2c_{ii}p^2_{\infty\omega}.$$

From this equation and those immediately preceding (2),

$$r_{\infty\omega} = (r_{12}r_{i,ii}/r_{1i}r_{2,ii})^{1/2}. \tag{4}$$

Taking the geometric mean of (3) and (4),

$$r_{\infty\omega} = (r_{12}r_{1,ii}r_{i2}r_{i,ii})^{1/4}/(r_{1i}r_{2,ii})^{1/2}. \tag{5}$$

Correlation corrected for attenuation, assuming all errors to be uncorrelated.

We might also combine (3) and (4) by taking their arithmetic mean, in which case,

$$r_{\infty\omega} = (r_{12}r_{i,ii} + r_{1,ii}r_{i2})^{1/2}/(2r_{1i}r_{2,ii})^{1/2}. \tag{6}$$

Alternate form of the correlation corrected for attenuation, assuming all errors to be uncorrelated.

Alternative formulas for the correlation corrected for attenuation. If we define the reliability of a test, as in Chapter I, by the relation,

$$R_1 = c_1^2\sigma_\infty^2/\sigma_1^2,$$

we find,

$$\sigma_\infty^2 = R_1\sigma_1^2/c_1^2.$$

Similarly,

$$\sigma_\omega^2 = R_2\sigma_2^2/c_2^2,$$

and from (1),

$$p^2_{\infty\omega} = p^2_{12}/c_1^2c_2^2,$$

so that,

$$r^2_{\infty\omega} = p^2_{12}/\sigma^2_1\sigma^2_2 R_1 R_2,$$

and,

$$r_{\infty\omega} = r_{12}/(R_1R_2)^{1/2}. \tag{7}$$

This formula will not in general be used as a substitute for (2) or (3) in practical computations, because of the difficulties that arise in evaluating R_1 and R_2. From it, however, we may deduce one important relationship. If $r_{\infty\omega}$ is equal to unity,

$$r_{12} = (R_1R_2)^{1/2}. \tag{8}$$

Upper limit of the obtained correlation between two fallible tests.

This limit is not absolute—it may be exceeded by chance—but for practical purposes it gives, as noted in the previous chapter, the upper limit of the ability of a fallible test to predict the score on another fallible test or criterion measure.

If $c_1/\sigma_1 = c_i/\sigma_i$ and $c_2/\sigma_2 = c_{ii}/\sigma_{ii}$, we have from equation (10) of Chapter II,

$$R_1 = R_i = r_{1i},$$
$$R_2 = R_{ii} = r_{2,ii},$$

and from (7),

$$r_{\infty\omega} = r_{12}/(r_{1i}r_{2,ii})^{1/2}. \tag{9}$$

By a line of reasoning exactly like that leading to (7) and (9), we obtain,

$$r_{\infty\omega} = r_{1,ii}/(r_{1i}r_{2,ii})^{1/2},$$
$$r_{\infty\omega} = r_{i2}/(r_{1i}r_{2,ii})^{1/2},$$
$$r_{\infty\omega} = r_{i,ii}/(r_{1i}r_{2,ii})^{1/2},$$

and averaging these equations and (9),

$$r_{\infty\omega} = (r_{12} + r_{1,ii} + r_{i2} + r_{i,ii})/(4(r_{1i}r_{2,ii})^{1/2}). \tag{10}$$

Correlation corrected for attenuation, when all the errors of measurement are uncorrelated and the two forms of each test are equally reliable.

Formulas (9) and (10) were given by Spearman in his original paper (1904 proof). He made the statement there,

"Should circumstances happen to render, say x_1, much more accurate than x_i, then the correlations involving x_1 will be considerably greater than those involving x_i. In such case, the

numerator of the above fraction must be formed by the geometrical mean instead of by the arithmetical mean; hereby the accidental errors of the respective observations cease to eliminate one another and therefore double their final influence; they also introduce an undue diminution of the fraction."*

Yule's proof of (5), cited by Brown (1909) and (1910 exp.), and Spearman (1910) was the first to show clearly the assumption that the errors of measurement are uncorrelated. It was based, as was his derivation of (3), on a factor pattern without the c's, assuming these to be equal for each test. With such a factor pattern, (10) can only be derived on the assumption of comparable tests of both abilities. The present derivation shows that the assumption of equal units is not necessary for the applicability of (3) and (5), and that the assumption of comparable tests is not basic to (10).

In practice, (3) and (5) will be the most useful for estimating true correlations, as they rest on only the single assumption that certain (or all) of the errors of measurement are uncorrelated with one another and with the true abilities.

Checks for the correctness of assumptions. There is no adequate check for the assumption that errors of measurement are uncorrelated. Brown (1909) and (1910 obj.), has proposed two. The first of these may be written in the present notation,

$$p_{1,ii} - p_{i2} = 0 \text{ (within the sampling error).} \qquad (11)$$

Now assuming that the errors of measurement *are* uncorrelated, we have from (1),

$$p_{1,ii} = c_1 c_{ii} p_{\infty\omega},$$
$$p_{i2} = c_i c_2 p_{\infty\omega},$$

whence we see that (11) holds only under the additional assumption that $c_1 c_{ii} = c_i c_2$, which was found not to be necessary in the derivation of (3) and (5). Brown's second check may be written,

$$r_{(x_1 - x_i)(x_2 - x_{ii})} = 0 \text{ (within the sampling error).} \qquad (12)$$

* In the above quotation, the notation here used has been substituted for that employed by Spearman. In this paper of 1904, no proof was given, but three years later (1907) another paper appeared, giving a lengthy and somewhat obscure proof of formula (5). In this paper Spearman noted that if the two sets of measurements in each case were equally reliable, formula (10) was to be preferred.

Expanding this we have,

$$(p_{12}+p_{i,ii}-p_{1,ii}-p_{i2})/(\sigma_{(x_1-x_i)}\sigma_{(x_2-x_{ii})})=0.$$

This expression will vanish with its numerator. Making the original assumption of zero correlation between errors of measurement again, we note from (1) that (11) will vanish only if $c_1c_2+c_ic_{ii}=c_1c_{ii}+c_ic_2$, a condition which again is not necessary to the derivation of (3) and (5). Neither of these checks is in general, therefore, a necessary condition for the applicability of either (3) or (5). But if the two forms of each test measure in the same units, both of them are applicable, and in fact,

$$p_{12}=p_{1,ii}=p_{i2}=p_{i,ii} \text{ (within the sampling errors)} \tag{13}$$

General check for the assumption that errors of measurement are uncorrelated, applicable when the two forms of each test are known to measure in the same units.

This check contains both of Brown's as special cases, and is founded on the same basic assumption of equality of units of measurement in the two forms of each test, that underlies Yule's simplified factor pattern, which Brown employed in deriving his checks. If (13) holds, (11) and (12) of necessity hold also.

As a check on the assumption that there are no correlated response errors, which is the essential condition for the applicability of (5), we have from (3) and (4),

$$r_{12}r_{i,ii}/r_{1,ii}r_{i2}=1. \tag{14}$$

Check for the assumption that response errors are uncorrelated.

This formula is a special case of the tetrad ratio.

CHAPTER V

The Theory of Factors

Problem and limitations. If a number of tests have been given to a group of individuals, we may desire to know the nature and relative contributions of the various basic mental abilities underlying the test performances. In previous chapters we have dealt with special aspects of this problem—aspects in which the nature of the abilities was known, in the sense that the fundamental factor pattern could be written down from a-priori considerations. In the more general cases now to be considered, the factor pattern cannot be determined beforehand. In fact, the central problem of the theory of factors is to find the simplest factor pattern consistent with a given set of data, and to effect an analysis of the variances of the different tests in terms of the factors of this pattern.

It is impossible to start with a set of test data and work backwards to a determination of the factor pattern. There are always an indefinite number of such patterns consistent with any set of data. For practical purposes, however, we may use the law of parsimony, and assume that the simplest of these is the "true" factor pattern. Then in ordinary situations we shall first assume some very simple pattern, derive the equations of consistency, and apply them to the data. If they fit, we shall go no further. If they do not, we shall postulate some slightly more complicated pattern, and repeat the process. As a matter of fact, however, it will be found that the same equations may often be used to test several hypothetical factor patterns. The present discussion will be limited to two of the simplest of these patterns.

The theory of two factors. This provides the simplest of all practical factor patterns. It assumes that among any set of tests not obviously similar, there will exist one single general factor common to them all, and that each test in addition will possess a specific factor independent of all other specific factors and of the general factor. The theory has been vigorously upheld by Spearman and his students for over a quarter-century. For four tests, the factor pattern may be written,

$$x_1 = c_1 g + s_1.$$
$$x_2 = c_2 g + s_2.$$
$$x_3 = c_3 g + s_3.$$
$$x_4 = c_4 g + s_4.$$

The value g is the general or common factor, and the s's are the specific factors. From this pattern we obtain,

$$p_{12} = c_1 c_2 \sigma_g^2.$$
$$p_{13} = c_1 c_3 \sigma_g^2.$$
$$p_{14} = c_1 c_4 \sigma_g^2.$$
$$p_{23} = c_2 c_3 \sigma_g^2.$$
$$p_{24} = c_2 c_4 \sigma_g^2.$$
$$p_{34} = c_3 c_4 \sigma_g^2.$$

from which we find

$$p_{12} p_{34} = p_{13} p_{24} = p_{14} p_{23} = c_1 c_2 c_3 c_4 \sigma_g^4. \tag{1}$$

Fundamental tetrad relation for the theory of two factors.

The equality of the three left-hand terms in (1) is the test for the consistency of a set of data with the theory of two factors. If we divide each of these terms by $\sigma_1 \sigma_2 \sigma_3 \sigma_4$, we obtain,

$$r_{12} r_{34} = r_{13} r_{24} = r_{14} r_{23}. \tag{2}$$

Tetrad relation in terms of correlation coefficients.

If we divide each of these in turn by $R_1 R_2 R_3 R_4$, we find,

$$r_{\infty\omega} r_{\gamma\eta} = r_{\infty\gamma} r_{\omega\eta} = r_{\infty\eta} r_{\omega\gamma}.$$

Tetrad relation in terms of correlations corrected for attenuation.

Application of the tetrad relation. It has been the custom of most investigators, following Spearman, to use the relation of (2) and derive therefrom the equations,

$$\left.\begin{aligned} r_{12} r_{34} - r_{13} r_{24} &= 0. \\ r_{12} r_{34} - r_{14} r_{23} &= 0. \\ r_{13} r_{24} - r_{14} r_{23} &= 0. \end{aligned}\right\} \tag{4}$$

Tetrad difference equations.

There are three more such equations which may be written by reversing the signs of these, but of the six, any two except a given one and its negative are sufficient to determine all the others.

There is no obvious reason except the historical one why the tetrad differences should be expressed in terms of raw correlation coefficients rather than in terms of covariances or of correlations corrected for attenuation. Any one of these sets of tetrad difference equations should equal zero according to the two-factor theory. But their sampling errors are different, as are also their numerical values when the theory of two factors does not hold. Thus if we write the first equation of (4) in terms of the covariances we have,

$$p_{12}p_{34}-p_{13}p_{24}=0.$$

In terms of the raw correlations this becomes,

$$r_{12}r_{34}-r_{13}r_{24}=(p_{12}p_{34}-p_{13}p_{24})/(\sigma_1\sigma_2\sigma_3\sigma_4)=0.$$

The sampling error now involves the errors not only of the covariances, but also of the standard deviations. In terms of the correlations corrected for attenuation, we find,

$$\begin{aligned}&r_{\infty\omega}r_{\gamma\eta}-r_{\infty\gamma}r_{\omega\eta}\\&=(p_{12}p_{34}-p_{13}p_{24})/(\sigma_1\sigma_2\sigma_3\sigma_4R_1R_2R_3R_4)=0.\end{aligned}$$

The sampling error here involves the errors of the reliability coefficients in addition to those of the covariances and standard deviations.

The tetrad ratio. There is another method of writing the tetrad relation which will give identical results, whether the data be given in terms of covariances, raw correlations, or correlations corrected for attenuation. This is the tetrad ratio. We may obtain from (1),

$$\left.\begin{aligned}p_{12}p_{34}/p_{13}p_{24}&=1.\\p_{12}p_{34}/p_{14}p_{23}&=1.\\p_{13}p_{24}/p_{14}p_{23}&=1.\end{aligned}\right\}\tag{5}$$

Tetrad ratio equations.

If we use raw correlation coefficients as our original data, we are simply dividing numerator and denominator simultaneously by the product of the four obtained standard deviations. These will always cancel. In a similar manner, if we take the correlations corrected for attenuation as the original data, we are again dividing numerator and denominator by the same value, in this case the product of the four obtained reliability coefficients. The numerical value of the tetrad ratio is the same, therefore, in each

case. This argument may be applied with equal force in the case of the true values for a theoretically infinite population, from which the given group may be assumed to be a random sample. There is only one value of the true tetrad ratio. Now no matter how the standard deviations and reliability coefficients of the sample may differ from those of the population, it is these obtained values which are substituted in the formula and which cancel. We see, therefore, that the sampling error of the tetrad ratio depends only on the errors in the covariances, no matter what coefficients we use as our basic data. This property should be of sufficient importance to cause investigators hereafter to employ the tetrad ratio in preference to the tetrad difference.

Experimental conditions for factor-theory studies. The theory of factors assumes that the correlations between tests are due to common underlying abilities. It is necessary, therefore, that the tests be given at intervals such that the response errors will not introduce general or group factors. If four tests of independent abilities be given at the same time, the common response errors will introduce a spurious general factor. It would seem, therefore, that each test would have to be given at a different time. But in extensive investigations this is rarely practical. If each test has two forms, all the Form A tests can be applied at one time and all the Form B tests at another. Then we may correlate the Form A score on any one of them with the Form B score or any other without introducing any correlated response errors. The factor pattern may then be written,

$$\begin{array}{ll} \text{Form A} & \text{Form B} \\ x_1 = c_1 g + k_1 s_1 + \delta_1. & x_i = c_i g + k_i s_1 + \delta_i. \\ x_2 = c_2 g + k_2 s_2 + \delta_2. & x_{ii} = c_{ii} g + k_{ii} s_2 + \delta_{ii}. \\ x_3 = c_3 g + k_3 s_3 + \delta_3. & x_{iii} = c_{iii} g + k_{iii} s_3 + \delta_{iii}. \\ x_4 = c_4 g + k_4 s_4 + \delta_4. & x_{iv} = c_{iv} g + k_{iv} s_4 + \delta_{iv}. \end{array} \Bigg\} \qquad (6)$$

The value g is the general factor, the s's are specific non-chance factors, and the δ's are errors of measurement. The δ's of the Form A tests are not assumed to be uncorrelated among themselves, nor are those of the Form B tests, but the Form A δ's *are* assumed to be uncorrelated with the Form B δ's. The c's and k's are constants, and $c_1/c_i = k_1/k_i$, $c_2/c_{ii} = k_2/k_{ii}$, $c_3/c_{iii} = k_3/k_{iii}$, and $c_4/c_{iv} = k_4/k_{iv}$, since the two forms of each test are supposed to differ only in their units of measurement and errors, but not in

their relative make-up in terms of underlying abilities. Then from this factor pattern we obtain,

$$\begin{aligned} p_{1,ii} &= c_1 c_{ii} \sigma_g^2. & p_{i2} &= c_i c_2 \sigma_g^2. \\ p_{1,iii} &= c_1 c_{iii} \sigma_g^2. & p_{i3} &= c_i c_3 \sigma_g^2. \\ p_{1,iv} &= c_1 c_{iv} \sigma_g^2. & p_{i4} &= c_i c_4 \sigma_g^2. \\ p_{2,iii} &= c_2 c_{iii} \sigma_g^2. & p_{ii,3} &= c_{ii} c_3 \sigma_g^2. \\ p_{2,iv} &= c_2 c_{iv} \sigma_g^2. & p_{ii,4} &= c_{ii} c_4 \sigma_g^2. \\ p_{3,iv} &= c_3 c_{iv} \sigma_g^2. & p_{iii,4} &= c_{iii} c_4 \sigma_g^2. \end{aligned}$$

From these equations we find,

$$\begin{aligned} & p_{1,ii} p_{i2} p_{3,iv} p_{iii,4} = p_{1,iii} p_{i3} p_{2,iv} p_{ii,4} \\ & = p_{1,iv} p_{i4} p_{2,iii} p_{ii,3} = c_1 c_i c_2 c_{ii} c_3 c_{iii} c_4 c_{iv} \sigma_g^8. \end{aligned} \tag{8}$$

Practical tetrad relation for the theory of two factors.

Using this relation we obtain,

$$\left.\begin{aligned} (p_{1,ii} p_{i2} p_{3,iv} p_{iii,4} / p_{1,iii} p_{i3} p_{2,iv} p_{ii,4})^{1/2} &= 1. \\ (p_{1,ii} p_{i2} p_{3,iv} p_{iii,4} / p_{1,iv} p_{i4} p_{2,iii} p_{ii,3})^{1/2} &= 1. \\ (p_{1,iii} p_{i3} p_{2,iv} p_{ii,4} / p_{1,iv} p_{i4} p_{2,iii} p_{ii,3})^{1/2} &= 1. \end{aligned}\right\} \tag{9}$$

Practical tetrad ratio equations.

There are three more equations of this sort which are the reciprocals of these. Only two are independent, as the third will always be the quotient (or its reciprocal) obtained by dividing one by another. Hence in practice it is only necessary to demonstrate that two of the tetrad ratios which are not reciprocals, are not significantly different from unity, in order to be able to assert the plausibility of the theory of two factors as an explanation of the data.

The correlation coefficients may replace the covariances in (9) without changing either the values or the standard errors of these equations. If we let the "true" scores underlying the two forms of tests one, two, three, and four be x_∞, x_ω, x_γ, and x_η respectively, we may compute $r_{\infty\omega}$, $r_{\infty\gamma}$, $r_{\infty\eta}$, $r_{\omega\gamma}$, $r_{\omega\eta}$, and $r_{\gamma\eta}$ by formula (3) of Chapter IV. If now, we go back to equations (5) and substitute these correlations corrected for attenuation, we obtain,

$$\left.\begin{aligned} r_{\infty\omega} r_{\gamma\eta} / r_{\infty\gamma} r_{\omega\eta} &= 1. \\ r_{\infty\omega} r_{\gamma\eta} / r_{\infty\eta} r_{\omega\gamma} &= 1. \\ r_{\infty\gamma} r_{\omega\eta} / r_{\infty\eta} r_{\omega\gamma} &= 1. \end{aligned}\right\} \tag{10}$$

Alternate form of the practical tetrad ratio equations.

The values yielded by (10) are identical with those of (9), as the denominators of all the correlations corrected for attenuation will cancel.

Analysis of variance in the theory of two factors. If equations (9) or (10) have been shown to hold for a particular set of data, we may determine the relative contributions of the general factor, the non-chance specific factor, and the error of measurement to the variance of any variable, provided the two forms of that variable are equally reliable. Consider x_1 and x_i. The assumption of equal reliability imposes the condition that $c_1/\sigma_1 = c_i/\sigma_i$, and $k_1/\sigma_1 = k_i/\sigma_i$, since $c_1/c_i = k_1/k_i$. From the first of these ratio equalities we obtain the important relation,

$$c_1^2\sigma_g^2/\sigma_1^2 = c_i^2\sigma_g^2/\sigma_i^2 = c_1c_i\sigma_g^2/\sigma_1\sigma_i. \tag{11}$$

From the first of equations (6), we find,

$$\begin{aligned} &c_1^2\sigma_g^2/\sigma_1^2 + k_1^2\sigma_{s_1}^2/\sigma_1^2 + \sigma_{\delta_1}^2/\sigma_1^2 \\ &= c_i^2\sigma_g^2/\sigma_i^2 + k_i^2\sigma_{s_1}^2/\sigma_i^2 + \sigma_{\delta_i}^2/\sigma_i^2 = 1. \end{aligned} \tag{12}$$

Analysis of variance of either form of variable one, when the two forms are equally reliable, and the theory of two factors holds.

Now from (7),

$$p_{1,ii}p_{i2} = c_1c_ic_2c_{ii}\sigma_g^4,$$
$$p_{1,iii}p_{i3} = c_1c_ic_3c_{iii}\sigma_g^4,$$
$$p_{2,iii}p_{ii,3} = c_2c_{ii}c_3c_{iii}\sigma_g^4,$$

and,

$$(p_{1,ii}p_{i2}p_{1,iii}p_{i3}/p_{2,iii}p_{ii,3})^{1/2} = c_1c_i\sigma_g^2.$$

From this equation and (11),

$$\begin{aligned} c_1^2\sigma_g^2/\sigma_1^2 = c_i^2\sigma_g^2/\sigma_1^2 &= (p_{1,ii}p_{i2}p_{1,iii}p_{i3}/p_{2,iii}p_{ii,3}\sigma_1^2\sigma_i^2)^{1/2} \\ &= (r_{1,ii}r_{i2}r_{1,iii}r_{i3}/r_{2,iii}r_{ii,3})^{1/2}. \end{aligned} \tag{13}$$

Proportion of the variance of either form of variable one due to the general factor, when the two forms are equally reliable, and the theory of two factors holds.

From the definition of the reliability coefficient and the assumption that both forms of the test are equally reliable, we obtain,

$$\sigma^2_{\delta_1}/\sigma^2_1 = \sigma^2_{\delta_i}/\sigma^2_i = 1 - r_{1i}. \tag{14}$$

Proportion of the variance of either form of variable one due to the error of measurement, when these forms are equally reliable.

This result is independent of the assumption that the theory of two factors holds. Finally,

$$k^2_1\sigma^2_{s_1}/\sigma^2_1 = k^2_i\sigma^2_{s_1}/\sigma^2_i$$

$$= r_{1i} - (r_{1,ii}r_{i2}r_{1,iii}r_{i3}/r_{2,iii}r_{ii,3})^{1/2}. \tag{15}$$

Proportion of the variance of either form of variable one due to the non-chance specific factor, when the two forms are equally reliable, and the theory of two factors holds.

An analysis similar to the above could be made for variable one, taking two and four or three and four as the reference variables, instead of two and three. The most accurate analysis would be made by using each of the three sets of reference variables in turn, and averaging the results. An analysis of the variances of two equally reliable forms of any of the other three tests could be made in the same manner.

The square root of any of these proportional contributions will be equal to the correlation between the contributing factor and the total variable. The proof in each case is similar to the proof that the square root of the reliability coefficient is equal to the correlation between the score and the underlying ability, as given in Chapter II for equation (12) of that Chapter.

The group-factor theory. According to this hypothesis, the correlations between tests are to be explained on the basis not only of one general factor and a specific factor in each variable, but in addition, of one or more independent group factors, which are common to some but not all of the observed variables. In the simplest case involving four tests, we shall assume that in addition to the general factor and the specific factors, there is a group factor common to two of the variables. We then have the following factor pattern,

$$\begin{array}{ll} \text{Form A} & \text{Form B} \\ x_1 = c_1 g + d_1 a + k_1 s_1 + \delta_1. & x_i = c_i g + d_i a + k_i s_1 + \delta_i. \\ x_2 = c_2 g + d_2 a + k_2 s_2 + \delta_2. & x_{ii} = c_{ii} g + d_{ii} a + k_{ii} s_2 + \delta_{ii}. \\ x_3 = c_3 g \qquad\quad + k_3 s_3 + \delta_3. & x_{iii} = c_{iii} g \qquad\quad + k_{iii} s_3 + \delta_{iii}. \\ x_4 = c_4 g \qquad\quad + k_4 s_4 + \delta_4. & x_{iv} = c_{iv} g \qquad\quad + k_{iv} s_4 + \delta_{iv}. \end{array} \quad (16)$$

The value g is the general factor, a is the group factor, the s's are specific non-chance factors, and the δ's are errors of measurement. We shall assume as before that all the different factors are uncorrelated except the δ's of Form A and the δ's of Form B, which may be correlated respectively among themselves but not with each other or with any of the other factors. The c's, d's and k's are constants. It is to be noted that $c_1/c_i = d_1/d_i = k_1/k_i$, etc., since the two forms of any test are supposed to differ only in their units of measurement and errors, and not in their relative proportions of underlying abilities. We may then write,

$$\left.\begin{array}{ll} p_{1,ii} = c_1 c_{ii} \sigma_g^2 + d_1 d_{ii} \sigma_a^2. & p_{i2} = c_i c_2 \sigma_g^2 + d_i d_2 \sigma_a^2. \\ p_{1,iii} = c_1 c_{iii} \sigma_g^2. & p_{i3} = c_i c_3 \sigma_g^2. \\ p_{1,iv} = c_1 c_{iv} \sigma_g^2. & p_{i4} = c_i c_4 \sigma_g^2. \\ p_{2,iii} = c_2 c_{iii} \sigma_g^2. & p_{ii,3} = c_{ii} c_3 \sigma_g^2. \\ p_{2,iv} = c_2 c_{iv} \sigma_g^2. & p_{ii,4} = c_{ii} c_4 \sigma_g^2. \\ p_{3,iv} = c_3 c_{iv} \sigma_g^2. & p_{iii,4} = c_{iii} c_4 \sigma_g^2. \end{array}\right\} \quad (17)$$

From these relations we obtain,

$$\left.\begin{array}{l} p_{1,iii} p_{i3} p_{2,iv} p_{ii,4} = p_{1,iv} p_{i4} p_{2,iii} p_{ii,3} = c_1 c_i c_2 c_{ii} c_3 c_{iii} c_4 c_{iv} \sigma_g^8. \\ p_{1,ii} p_{i2} p_{3,iv} p_{iii,4} = c_1 c_i c_2 c_{ii} c_3 c_{iii} c_4 c_{iv} \sigma_g^8 + \text{terms containing } \sigma_a^2. \end{array}\right\} \quad (18)$$

Tetrad relation for the single group-factor theory.

If the group factor had been in x_3, x_{iii}, x_4, and x_{iv} instead of in x_1, x_i, x_2, and x_{ii}, the last pair of equations (17) would have contained the σ_a^2 term instead of the first pair. In this case, however, (18) would remain unchanged. Hence we see that any factor-pattern tests derived from (18) can only establish the fact that a group factor exists either in variables one and two, or in three and four, or in each of these pairs. From (18),

$$\left.\begin{array}{l} (p_{1,ii} p_{i2} p_{3,iv} p_{iii,4} / p_{1,iii} p_{i3} p_{2,iv} p_{ii,4})^{1/2} \\ = (p_{1,ii} p_{i2} p_{3,iv} p_{iii,4} / p_{1,iv} p_{i4} p_{2,iii} p_{ii,3})^{1/2} \neq 1, \\ (p_{1,iii} p_{i3} p_{2,iv} p_{ii,4} / p_{1,iv} p_{i4} p_{2,iii} p_{ii,3})^{1/2} = 1, \text{ or} \\ r_{\infty\omega} r_{\gamma\eta} / r_{\infty\gamma} r_{\omega\eta} = r_{\infty\omega} r_{\gamma\eta} / r_{\infty\eta} r_{\omega\gamma} \neq 1, \\ r_{\infty\gamma} r_{\omega\eta} / r_{\infty\eta} r_{\omega\gamma} = 1. \end{array}\right\} \quad (19)$$

Tetrad ratio equations for the single group-factor theory.

If we divide the second tetrad ratio of (19) by the first, their quotient will be the third. Hence the substantial equality of the first two is demonstrated at once when the third is shown to be equal (within its sampling error) to unity. The covariances of the two pairs of variables in either or both of which the group factor may lie, are found in the numerators of the fractions which are not equal to unity. It is possible to form three more tetrad ratio equations which will be the reciprocals of those given in (19), but these are not needed in any further analysis.

Analysis of variance in the single group-factor theory. The demonstration that in a set of three tetrad ratios, one is substantially equal to unity and one other is not, is sufficient to establish the fact that the theory of two factors does not hold, that the single group-factor theory may, and that if it does, the group factor lies in one or the other or both of two particular pairs of variables. But that is all it does establish. If we take each of the pairs of variables suspected of containing a group factor with other pairs, we may be able to discover in which pair it actually resides. Sometimes such an analysis is unnecessary, as the variables containing the group factor can be picked out simply by an examination of the nature of the tests. If we know that of four variables, only one particular pair contains a group factor, it is possible to effect a partial analysis of variance of the different variables. Consider variable one again, when the factor pattern is known to be that given in (16). We must assume as before that the two forms of test one are equally reliable, so that $c_1/\sigma_1 = c_i/\sigma_i$, and equation (11) still applies. From the first of equations (16),

$$
\begin{aligned}
& c_1^2\sigma_g^2/\sigma_1^2 + d_1^2\sigma_a^2/\sigma_1^2 + k_1^2\sigma_{s_1}^2/\sigma_1^2 + \sigma_{\delta_1}^2/\sigma_1^2 \\
& = c_i^2\sigma_g^2/\sigma_i^2 + d_i^2\sigma_a^2/\sigma_i^2 + k_i^2\sigma_{s_1}^2/\sigma_i^2 + \sigma_{\delta_i}^2/\sigma_i^2 = 1. \qquad (20)
\end{aligned}
$$

Analysis of variance of either form of variable one, when these forms are equally reliable, and the single group-factor theory holds.

Now from (17),

$$
\begin{aligned}
p_{1,iii}p_{i3} &= c_1 c_i c_3 c_{iii}\sigma_g^4, \\
p_{1,iv}p_{i4} &= c_1 c_i c_4 c_{iv}\sigma_g^4, \\
p_{3,iv}p_{iii,4} &= c_3 c_{iii} c_4 c_{iv}\sigma_g^4,
\end{aligned}
$$

and,

$$(p_{1,iii}p_{i3}p_{1,iv}p_{i4}/p_{3,iv}p_{iii,4})^{1/2} = c_1 c_i \sigma_g^2.$$

From this equation and (11),

$$c_1^2\sigma_g^2/\sigma_1^2 = c_i^2\sigma_g^2/\sigma_i^2 = (p_{1,iii}p_{i3}p_{1,iv}p_{i4}/p_{3,iv}p_{iii,4}\sigma_1^2\sigma_i^2)^{1/2}$$
$$= (r_{1,iii}r_{i3}r_{1,iv}r_{i4}/r_{3,iv}r_{iii,4})^{1/2}. \quad (21)$$

Proportion of the variance of either form of variable one due to the general factor, when there is a single group factor in variables one and two, and the two forms of test one are equally reliable.

This equation is similar to (13), but subject to the limitation that the reference variables *must* be those which do not contain the group factor. Equation (14) still applies, giving the proportion of the variance due to errors of measurement. This is as far as it is possible to proceed with the analysis when there are only four variables. The difference between the proportion of the variance due to the general factor and the proportion due to the error of measurement will equal the proportion due to the group factor and the non-chance specific factor taken together. These last two proportions cannot be separated on the basis of data from four variables.

The variance of variable two can be partially analyzed in similar fashion. Those of variables three and four can be analyzed completely, as described for the case where the theory of two factors holds. Variables one and two should not be taken together as the reference variables. Hence, in analyzing the variance of variable three we could take one and four or two and four; and in analyzing variable four, one and three or two and three, as the reference variables, but in neither case could we take one and two.

Analysis of covariance. Consider the covariance of x_1 and x_{ii}, and of x_i and x_2. From (17),

$$p_{1,iii}p_{ii,3} = c_1 c_{ii} c_3 c_{iii} \sigma_g^4,$$
$$p_{1,iv}p_{ii,4} = c_1 c_{ii} c_4 c_{iv} \sigma_g^4,$$
$$p_{3,iv}p_{iii,4} = c_3 c_{iii} c_4 c_{iv} \sigma_g^4,$$

and,

$$(p_{1,iii}p_{ii,3}p_{1,iv}p_{ii,4}/p_{3,iv}p_{iii,4})^{1/2} = c_1 c_{ii} \sigma_g^2 = \gamma_{1,ii}, \text{ say.} \quad (22)$$

Amount of the covariance of x_1 and x_{ii} due to the general factor.

$$p_{i3}p_{2,iii} = c_i c_2 c_3 c_{iii} \sigma_g^4,$$
$$p_{i4}p_{2,iv} = c_i c_2 c_4 c_{iv} \sigma_g^4,$$
$$p_{3,iv}p_{iii,4} = c_3 c_{iii} c_4 c_{iv} \sigma_g^4,$$

and,

$$(p_{i3}p_{2,iii}p_{i4}p_{2,iv}/p_{3,iv}p_{iii,4})^{1/2} = c_i c_2 \sigma_g^2 = \gamma_{i,2}. \tag{23}$$

Amount of the covariance of x_i and x_2 due to the general factor.

Then from the first of equations (17),

$$d_1 d_{ii} \sigma_a^2 = p_{1,ii} - \gamma_{1,ii} = \alpha_{1,ii}, \text{ say.} \tag{24}$$

Amount of the covariance of x_1 and x_{ii} due to the group factor.

Also,

$$d_i d_2 \sigma_a^2 = p_{i2} - \gamma_{i,2} = \alpha_{i,2}. \tag{25}$$

Amount of covariance of x_i and x_2 due to the group factor. All the other covariances, as may be seen from (17), are due entirely to the general factor.*

Analysis of variance with five variables. If we have five variables with a group factor running through three of them, it is possible to make a complete analysis of variance. The presence of such a group factor may be established by equations (19), taking each of the three possible pairs of variables suspected of containing it, with a number of pairs of other variables which among themselves conform to the theory of two factors. The factor pattern will be exactly like (16), with the single additional pair of relations,

$$x_5 = c_5 g + d_5 a + k_5 s_5 + \delta_5. \qquad x_v = c_v g + d_v a + k_v s_5 + \delta_v.$$

Then in addition to (17) we will have,

$$\left.\begin{array}{ll}
p_{1,v} = c_1 c_v \sigma_g^2 + d_1 d_v \sigma_a^2. & p_{i5} = c_i c_5 \sigma_g^2 + d_i d_5 \sigma_a^2. \\
p_{2v} = c_2 c_v \sigma_g^2 + d_2 d_v \sigma_a^2. & p_{ii,5} = c_{ii} c_5 \sigma_g^2 + d_{ii} d_5 \sigma_a^2. \\
p_{3v} = c_3 c_v \sigma_g^2. & p_{iii,5} = c_{iii} c_5 \sigma_g^2. \\
p_{4v} = c_4 c_v \sigma_g^2. & p_{iv,5} = c_{iv} c_5 \sigma_g^2.
\end{array}\right\} \tag{26}$$

By an argument similar to that leading to (22), (23), (24), and (25), we have,

---

* I am indebted to my colleagues, Mr. Jack W. Dunlap and Dr. Irving Lorge, for suggesting the above analysis of covariance, and the analysis of variance in the case of five variables, which follows.

$$(p_{1,iii}p_{3v}p_{1,iv}p_{4v}/p_{3,iv}p_{iii,4})^{1/2} = c_1c_v\sigma_g^2 = \gamma_{1,v}.$$
$$(p_{i3}p_{iii,5}p_{i4}p_{iv,5}/p_{3,iv}p_{iii,4})^{1/2} = c_ic_5\sigma_g^2 = \gamma_{i,5}.$$
$$(p_{2,iii}p_{3v}p_{2,iv}p_{4v}/p_{3,iv}p_{iii,4})^{1/2} = c_2c_v\sigma_g^2 = \gamma_{2,v}.$$
$$(p_{ii,3}p_{iii,5}p_{ii,4}p_{iv,5}/p_{3,iv}p_{iii,4})^{1/2} = c_{ii}c_5\sigma_g^2 = \gamma_{ii,5}.$$

Then,

$$d_1d_v\sigma_a^2 = p_{1v} - \gamma_{1,v} = \alpha_{1,v}. \tag{27}$$
$$d_id_5\sigma_a^2 = p_{i,5} - \gamma_{i,5} = \alpha_{i,5}. \tag{28}$$
$$d_2d_v\sigma_a^2 = p_{2,v} - \gamma_{2,v} = \alpha_{2,v}. \tag{29}$$
$$d_{ii}d_5\sigma_a^2 = p_{ii,5} - \gamma_{ii,5} = \alpha_{ii,5}. \tag{30}$$

The six equations, (24), (25), (27), (28), (29) and (30), form a system which may be solved for $d_1d_i\sigma_a^2$, $d_2d_{ii}\sigma_a^2$, and $d_5d_v\sigma_a^2$. If we assume that the two forms of each of the tests are equally reliable, we find that $c_1/\sigma_1 = c_i/\sigma_i$, $d_1/\sigma_1 = d_i/\sigma_i$, $k_1/\sigma_1 = k_i/\sigma_i$, and that similar relations hold for variables two and five, since $c_1/c_i = d_1/d_i = k_1/k_i$, etc. From the first of these ratio equations we obtain equation (11) again, and from the second,

$$d_1^2\sigma_a^2/\sigma_1^2 = d_i^2\sigma_a^2/\sigma_i^2 = d_1d_i\sigma_a^2/\sigma_1\sigma_i. \tag{31}$$

From (24), (25), (27), (28), (29), and (30),

$$(\alpha_{1,ii}\alpha_{i,2}\alpha_{1,v}\alpha_{i,5}/\alpha_{2,v}\alpha_{ii,5})^{1/2} = d_1d_i\sigma_a^2,$$
$$(\alpha_{1,ii}\alpha_{i,2}\alpha_{2,v}\alpha_{ii,5}/\alpha_{1,v}\alpha_{i,5})^{1/2} = d_2d_{ii}\sigma_a^2,$$
$$(\alpha_{1,v}\alpha_{i,5}\alpha_{2,v}\alpha_{ii,5}/\alpha_{1,ii}\alpha_{i,2})^{1/2} = d_5d_v\sigma_a^2,$$

and from (31),

$$d_1^2\sigma_a^2/\sigma_1^2 = d_i^2\sigma_a^2/\sigma_i^2 = d_1d_i\sigma_a^2/\sigma_1\sigma_i$$
$$= (\alpha_{1,ii}\alpha_{i,2}\alpha_{1,v}\alpha_{i,5}/\alpha_{2,v}\alpha_{ii,5}\sigma_1^2\alpha_i^2)^{1/2} \tag{32}$$

Proportion of the variance of either form of variable one due to the group factor, when there is a single group factor in variables one, two, and five, and the two forms of each of these tests are equally reliable.

Similar equations can of course be written for variables two and five. Equations (14) and (21) give the proportions of the variance of variable one due to the chance factor and the general factor respectively, as before. We have finally,

$$k_1^2\sigma_{s_1}^2/\sigma_1^2 = k_i^2\sigma_{s_1}^2/\sigma_i^2 = k_1k_i\sigma_{s_1}^2/\sigma_1\sigma_i$$
$$= r_{1i} - (r_{1,iii}r_{i3}r_{1,iv}r_{i4}/r_{3,iv}r_{iii,4})^{1/2}$$
$$- (\alpha_{1,ii}\alpha_{i,2}\alpha_{1,v}\alpha_{i,5}/\alpha_{2,v}\alpha_{ii,5}\sigma_1^2\sigma_i^2)^{1/2}. \tag{33}$$

Proportion of the variance of either form of variable one due to the group factor, when there is a single group factor in variables one, two, and five, and the two forms of each of these tests are equally reliable.

Analysis of "true" variance. The whole problem of the analysis of variance may be approached in another manner by taking the estimated "true" variance as unity instead of the obtained variance. If the two forms of a test are equally reliable, we have from the definition of the reliability coefficient,

$$c_1^2\sigma_\infty^2 = \sigma_1^2 r_{1i}. \tag{34}$$

Estimated "true" variance of variable one, measured in the units of x_1.

The constant c_1^2 indicates merely that σ_∞^2 is measured in the units of Form A of variable one. An alternative analysis of the variance of x_1 or x_i, when these are equally reliable, may be made by analyzing the variance of x_∞ and multiplying the values so obtained by r_{1i}. In the more complicated cases involving five variables, this method of analysis is simpler than the one previously outlined.

Let the portions of variables one, two, three, and four which are not chance be denoted x_∞, x_ω, x_γ, and x_η. The factor pattern may then be written,

$$\left.\begin{aligned} x_\infty &= c_1 g + s_1. \\ x_\omega &= c_2 g + s_2. \\ x_\gamma &= c_3 g + s_3. \\ x_\eta &= c_4 g + s_4. \end{aligned}\right\} \tag{35}$$

We then define the values,

$$\left.\begin{aligned} \alpha_1 &= c_1\sigma_g/\sigma_\infty. \\ \alpha_2 &= c_2\sigma_g/\sigma_\omega. \\ \alpha_3 &= c_3\sigma_g/\sigma_\gamma. \\ \alpha_4 &= c_4\sigma_g/\sigma_\eta. \end{aligned}\right\} \tag{36}$$

From (35) and (36) we find,

$$\left.\begin{aligned} r_{\infty\omega} &= \alpha_1\alpha_2. & r_{\omega\gamma} &= \alpha_2\alpha_3. \\ r_{\infty\gamma} &= \alpha_1\alpha_3. & r_{\omega\eta} &= \alpha_2\alpha_4. \\ r_{\infty\eta} &= \alpha_1\alpha_4. & r_{\gamma\eta} &= \alpha_3\alpha_4. \end{aligned}\right\} \tag{37}$$

Then,

$$r_{\infty\omega}r_{\gamma\eta} = r_{\infty\gamma}r_{\omega\eta} = r_{\infty\eta}r_{\omega\gamma} = \alpha_1\alpha_2\alpha_3\alpha_4. \tag{38}$$

This is the tetrad relation, and it has already been proved that the tetrad ratio in this case is the same as that obtained from the covariances or correlation coefficients.

The analysis of variance in this situation is comparatively simple. From (36) and (37),

$$c_1^2\sigma_g^2/\sigma_\infty^2 = \alpha_1^2 = r_{\infty\omega}r_{\infty\gamma}/r_{\omega\gamma}$$
$$= r_{\infty\omega}r_{\infty\eta}/r_{\omega\eta} = r_{\infty\gamma}r_{\infty\eta}/r_{\gamma\eta}. \tag{39}$$

Proportion of the "true" variance of variable one due to the general factor, when the theory of two factors holds.

This formula has been used by Cureton and Dunlap (1930), and called the triad by them.

$$\sigma_{s_1}^2/\sigma_\infty^2 = 1 - r_{\infty\omega}r_{\infty\gamma}/r_{\omega\gamma}$$
$$= 1 - r_{\infty\omega}r_{\infty\eta}/r_{\omega\eta} = 1 - r_{\infty\gamma}r_{\infty\eta}/r_{\gamma\eta}. \tag{40}$$

Proportion of the "true" variance of variable one due to the specific factor, when the theory of two factors holds.

Note that the term, "specific factor" as used here, means the non-chance specific factor. The error of measurement or chance specific factor is eliminated from consideration entirely by the definition of the "true" variance and the use of correlations corrected for attenuation.

For the single group-factor theory we have the factor pattern,

$$\left.\begin{aligned} x_\infty &= c_1g + d_1a + s_1. \\ x_\omega &= c_2g + d_2a + s_2. \\ x_\gamma &= c_3g \qquad\quad + s_3. \\ x_\eta &= c_4g \qquad\quad + s_4. \\ x_\Theta &= c_5g + d_5a + s_5. \end{aligned}\right\} \tag{41}$$

If we define,

$$\beta_1 = d_1\sigma_a/\sigma_\infty,$$
$$\beta_2 = d_2\sigma_a/\sigma_\omega,$$
$$\beta_5 = d_5\sigma_a/\sigma_\Theta,$$

we obtain,

$$\left.\begin{aligned} r_{\infty\omega} &= \alpha_1\alpha_2 + \beta_1\beta_2. & r_{\omega\eta} &= \alpha_2\alpha_4. \\ r_{\infty\gamma} &= \alpha_1\alpha_3. & r_{\omega\Theta} &= \alpha_2\alpha_5 + \beta_2\beta_5. \\ r_{\infty\eta} &= \alpha_1\alpha_4. & r_{\gamma\eta} &= \alpha_3\alpha_4. \\ r_{\infty\Theta} &= \alpha_1\alpha_5 + \beta_1\beta_5. & r_{\gamma\Theta} &= \alpha_3\alpha_5. \\ r_{\omega\gamma} &= \alpha_2\alpha_3. & r_{\eta\Theta} &= \alpha_4\alpha_5. \end{aligned}\right\} \tag{42}$$

Then,

$$c_1^2\sigma_g^2/\sigma_\infty^2 = \alpha_1^2 = r_{\infty\gamma}r_{\infty\eta}/r_{\gamma\eta}. \tag{43}$$

Proportion of the "true" variance of variable one due to the general factor, when the single group-factor theory holds.

This is the particular one of equations (39) which does not take x_ω as one of the reference variables.

Analyzing now the correlation $r_{\infty\omega}$,

$$\alpha_1\alpha_2 = r_{\infty\gamma}r_{\omega\eta}/r_{\gamma\eta}. \tag{44}$$

Amount of the "true" correlation between variables one and two due to the general factor.

$$\beta_1\beta_2 = r_{\infty\omega} - r_{\infty\gamma}r_{\omega\eta}/r_{\gamma\eta}. \tag{45}$$

Amount of the "true" correlation between variables one and two due to the group factor, when the single group-factor theory holds.

By a similar line of reasoning,

$$\beta_1\beta_5 = r_{\infty\Theta} - r_{\infty\gamma}r_{\eta\Theta}/r_{\gamma\eta},$$

and,

$$\beta_2\beta_5 = r_{\omega\Theta} - r_{\gamma\Theta}r_{\omega\eta}/r_{\gamma\eta}.$$

From these equations and (45),

$$d_1^2\sigma_a^2/\sigma_\infty^2 = \beta_1^2 = (\beta_1\beta_2)(\beta_1\beta_5)/(\beta_2\beta_5). \tag{46}$$

Proportion of the "true" variance of variable one due to the group factor, when the single group-factor theory holds.

Finally,

$$\sigma_{s_1}^2/\sigma_\infty^2 = 1 - \alpha_1^2 - \beta_1^2. \tag{47}$$

Proportion of the "true" variance of variable one due to the specific factor, when the single group-factor theory holds.

Variables two and five can be analyzed in similar fashion, and variables three and four can be analyzed as was variable one for the two-factor-theory case, taking care not to use variables one and two or one and five or two and five as the reference variables.

CHAPTER VI

Summary of Important Formulas, with their Standard Errors

Assumptions involved in standard error derivations. The principal assumptions involved in the derivation of the standard errors given hereafter are:

1. That all samples are drawn from populations normally distributed with respect to all the variables measured.

2. That all samples are drawn from populations in which the regressions of all the variables on one another are linear.

3. That all samples are sufficiently large so that higher powers of the sampling errors are small in comparison with first powers, and may be neglected.

The sampling errors of correlation functions depend essentially on the simultaneous error-distribution of the variances and covariances of the system. This distribution has been determined by Wishart (1928), who provides a table of its moments up to the eighth order and four variables, in terms of the variances and correlations of the sampled population. Since we have assumed a large sample, we may replace population values by sample values. Furthermore, we may replace the value $(N-1)/N^2$ by $1/N$, N being the total frequency of the sample. Then, expressing our results in terms of variances and covariances instead of in terms of variances and correlation coefficients, we have from Wishart's table of moments,

$$\left.\begin{aligned}
N\sigma^2_{\sigma_1^2} &= 2\sigma_1^4.\\
NP_{\sigma_1^2\sigma_2^2} &= 2p_{12}^2.\\
N\sigma^2_{p_{12}} &= \sigma_1^2\sigma_2^2+p_{12}^2.\\
NP_{\sigma_1^2 p_{12}} &= 2\sigma_1^2 p_{12}.\\
NP_{\sigma_1^2 p_{23}} &= 2p_{12}p_{13}.\\
NP_{p_{12}p_{13}} &= \sigma_1^2 p_{23}+p_{12}p_{13}.\\
NP_{p_{12}p_{34}} &= p_{13}p_{24}+p_{14}p_{23}.
\end{aligned}\right\} \qquad (1)$$

The large σ^2 and the capital P represent the sampling variance and covariance respectively.

Most of the important functions discussed in the previous chapters are in a form consisting of a series of products and quotients of variances and covariances. With the assumptions stated above, the standard error of any such function may be determined to a first approximation by taking its differential or logarithmic differential, squaring, summing for all samples, dividing by the number of samples, and substituting the values of sampling variances and covariances from (1).

There is one other approximation that enters into certain formulas. If we know that the sampling error is F^2, say, is Δ, we may wish to know the corresponding error in F. Now $(F^2+\Delta)^{1/2} = F+\Delta/2F-\Delta^2/8F^3+\ldots$, and if F is large as compared to Δ, the term $\Delta^2/8F^3$, together with all subsequent terms in the expansion, will be negligible in comparison with $\Delta/2F$. Then if δ is the error in F, $\delta=\Delta/2F$, to a first approximation.

Notation. In order to facilitate the work of derivation, a new system of notation has been introduced. Its relation to the system used in the previous chapters will be apparent at once from the table following.

Form A		Form B	
New	Old	New	Old
1	1	3	i
4	2	2	ii
5	3	7	iii
8	4	6	iv
9	5	11	v
12	6	10	vi

This system was designed so that a series of products such as $p_{1,ii}p_{i2}p_{3,iv}p_{iii,4}\ldots$ could be represented by $p_{12}p_{34}p_{56}p_{78}\ldots$

In the following paragraphs, the more important formulas of the previous chapters are given again in the new notation. In many cases two such formulas are of the same form algebraically. In each case the original uses of the formula are given, together with the chapter and formula numbers under which it has previously appeared. Its standard error, to the degree of approximation stated above, is presented. The algebra involved in the derivation of these standard errors is too lengthy to be included here.*

* A copy of these derivations is on file at the Library of Teachers College, Columbia University.

1.

$$F = r_{13}^{1/2}. \tag{2}$$

Index of reliability determined from two equally reliable forms (II, (12) and (10)).

$$\sigma_F = (1 - r_{13}^2)/2(r_{13}N)^{1/2}. \tag{3}$$

2.

$$F = nr_{13}/(1 + (n-1)r_{13}). \tag{4}$$

Spearman-Brown formula (II, (28)).

$$\sigma_F = (n - nr_{13}^2)/[N^{1/2}(1 + (n-1)r_{13})^2]. \tag{5}$$

This formula was first given by Shen (1924). If $n = 2$, we have the special case,

$$F' = 2r_{13}/(1 + r_{13}). \tag{6}$$

Reliability of the sum of two comparable forms of a test (II, (15)).

$$\sigma_{F'} = (2 - 2F')/N^{1/2}. \tag{7}$$

3.

$$F = \sigma_1 - \sigma_3. \tag{8}$$

Partial check on the comparability of two forms of a test. See discussion immediately preceding II, (27).

$$\sigma_F = (\sigma_1^2 + \sigma_3^2 - 2r_{13}^2\sigma_1\sigma_3)^{1/2}/(2N)^{1/2}. \tag{9}$$

4.

$$F = p_{12} - p_{13} \tag{10}$$

$$F' = p_{12} - p_{34} \tag{11}$$

Test for equality of basic units of measurement of three or more forms of a test. See discussion immediately preceding II, (22).

Check on assumption that errors of measurement are uncorrelated, applicable when the two forms of each test are known to measure in the same basic units (IV, (13)).

$$\sigma_F = (\sigma_1^2(\sigma_2^2 + \sigma_3^2 - 2p_{23}) + p_{12}^2 + p_{13}^2 - 2p_{12}p_{13})^{1/2}/N^{1/2}. \tag{12}$$

$$\sigma_{F'} = (\sigma_1^2\sigma_2^2 + \sigma_3^2\sigma_4^2 + p_{12}^2 + p_{34}^2 - 2p_{13}p_{24} - 2p_{14}p_{23})^{1/2}/N^{1/2}. \tag{13}$$

5.

$$F = \sigma_{p_{12}} = (\sigma_1^2\sigma_2^2 + p_{12}^2)^{1/2}/N^{1/2}. \tag{14}$$

Test for equality of basic units of measurement of several forms of a test (at least 6 or 7). See discussion immediately preceding II, (22).

In this formula, p_{12} is the average of all the covariances of the several forms, and σ_1 and σ_2 are the lower and upper quartile values of the distribution of obtained standard deviations. The value of F is to be compared with $\sigma_{p_{jk}}$, the standard deviation of the distribution of observed covariances. If a value of F as great or greater than $\sigma_{p_{jk}}$ could reasonably arise by chance (as judged by σ_F), then the several forms may be assumed all to be measuring in the same units.

$$\sigma_F = (\sigma_1^4\sigma_2^4 + p_{12}^4 + 6p_{12}^2\sigma_1^2\sigma_2^2)^{1/2}/(N^2\sigma_1^2\sigma_2^2 + N^2p_{12}^2)^{1/2}. \tag{15}$$

6.

$$F = r_{12}r_{13}/r_{23} = p_{12}p_{13}/p_{23}\sigma_1^2. \tag{16}$$

Reliability of Form A determined from a knowledge of three tests of the same ability (II, (17), (18), (19)).

Coefficient of practical validity (III, (6)).

Square of correlation between test 1 and g, or proportion of variance of test one due to g. See Spearman (1927), Appendix p. xvi, Kelley (1928), p. 41, and Dunlap (1931).

$$\sigma_F = (F/N^{1/2})(1/r_{12}^2 + 1/r_{13}^2 + 1/r_{23}^2 + 4F + 2/F - 2r_{13}/r_{12}r_{23} - 2r_{12}/r_{13}r_{23} - 5)^{1/2}. \tag{17}$$

A formula for the standard error of this function was given by Kelley (1928), pp. 40-41, based on the fundamental formulas of Filon and Pearson (1898). In the form there given it is much longer than (17).

7.

$$F = (r_{12}r_{14}r_{23}r_{34})^{1/4}/(r_{13}r_{24})^{1/2} = (p_{12}p_{14}p_{23}p_{34})^{1/4}/(p_{13}p_{14})^{1/2}. \tag{18}$$

Coefficient of fundamental validity (III, (4)).

Correlation corrected for attenuation, when all the errors of measurement are uncorrelated (IV, (5)).

$$\begin{aligned}\sigma_F = (F/4N^{1/2})[&1/r_{12}^2 + 1/r_{14}^2 + 1/r_{23}^2 + 1/r_{34}^2 + 4/r_{13}^2 + 4/r_{24}^2 \\ &+ 2(r_{13}r_{24} + r_{14}r_{23})/r_{12}r_{34} + 2(r_{12}r_{34} + r_{13}r_{24})/r_{14}r_{23} \\ &+ 8(r_{12}r_{34} + r_{14}r_{23})/r_{13}r_{24} + 2(r_{13}/r_{12}r_{23} + r_{13}/r_{14}r_{34} \\ &+ r_{24}/r_{12}r_{14} + r_{24}/r_{23}r_{34}) - 4(r_{12}/r_{13}r_{23} + r_{12}/r_{14}r_{24} \\ &+ r_{14}/r_{12}r_{24} + r_{14}/r_{13}r_{34} + r_{23}/r_{12}r_{13} + r_{23}/r_{24}r_{34} \\ &+ r_{34}/r_{13}r_{14} + r_{34}/r_{23}r_{24}) - 12]^{1/2}.\end{aligned} \tag{19}$$

A simpler formula for the standard error of this function was given by Kelley (1923), p. 210, based on the additional assumption that $r_{12}=r_{14}=r_{23}=r_{34}=r$, say. This assumption is justified whenever it can be shown that the two forms of each test are equally reliable. In this case,

$$F'=r/(r_{13}r_{24})^{1/2}. \tag{20}$$

$$\sigma_{F'}=(F'/2N^{1/2})(1/r_{13}^2+1/r_{24}^2+1/r^2+4F'^2+1/F'^2 +r_{13}/r^2+r_{24}/r^2-4/r_{13}-4/r_{24}-2)^{1/2}. \tag{21}$$

This formula is identical with Kelley's. In computation, r is to be taken as $(r_{12}r_{14}r_{23}r_{34})^{1/4}$.

8.

$$F=r_{12}r_{34}/r_{13}r_{24}=p_{12}p_{34}/p_{13}p_{24}. \tag{22}$$

Tetrad ratio (V, (5)).

Partial check for equivalence of tests and criterion measures (III, (5), letting subscripts 1, 2, i, ii = 1, 2, 3, 4, respectively).

Check for the assumption that response errors are uncorrelated (IV, (14)).

$$\sigma_F=(F/N^{1/2})[1/r_{12}^2+1/r_{34}^2+1/r_{13}^2+1/r_{24}^2+2(1/F+F +r_{14}r_{23}/r_{12}r_{34}+r_{14}r_{23}/r_{13}r_{24}-r_{23}/r_{12}r_{13} -r_{23}/r_{24}r_{34}-r_{14}/r_{12}r_{24}-r_{14}/r_{13}r_{34})-4]^{1/2}. \tag{23}$$

9.

$$F=(r_{12}r_{34}/r_{13}r_{24})^{1/2}=(p_{12}p_{34}/p_{13}p_{24})^{1/2}. \tag{24}$$

Correlation corrected for attenuation (IV, (3) and (4), the latter by letting the subscripts 1, 2, i, ii, equal 1, 2, 3, 4, respectively).

$$\sigma_F=(F/2N^{1/2})[1/r_{12}^2+1/r_{34}^2+1/r_{13}^2+1/r_{24}^2+2(1/F^2+F^2 +r_{14}r_{23}/r_{12}r_{34}+r_{14}r_{23}/r_{13}r_{24}-r_{23}/r_{12}r_{13} -r_{23}/r_{24}r_{34}-r_{14}/r_{12}r_{24}-r_{14}/r_{13}r_{34})-4]^{1/2}. \tag{25}$$

If the two forms of each test are equally reliable, we may assume that $r_{12}=r_{34}=r$, and $r_{14}=r_{23}=r'$. In this case,

$$F'=r/(r_{13}r_{24})^{1/2}. \tag{26}$$

$$\sigma_{F'}=(F'/2N^{1/2})[1/r_{13}^2+1/r_{24}^2+2(F'^2+1/F'^2+1/r^2 +r'^2/r^2+r'^2/r_{13}r_{24}-2r'/r_{13}-2r'/r_{24})-4]^{1/2}. \tag{27}$$

The values of r and r′ for purposes of computation should be taken as $(r_{12}r_{34})^{1/2}$ and $(r_{14}r_{23})^{1/2}$ respectively.

10. The standard errors of many formulas depend upon the sampling variances and covariances of correlations corrected for attenuation. The sampling variance is simply the square of the standard error, which may be obtained from (25) or (27).

$$r_{\infty\omega} = (r_{12}r_{34}/r_{13}r_{24})^{1/2}.$$

$$r_{\infty\gamma} = (r_{17}r_{35}/r_{13}r_{57})^{1/2}.$$

$$r_{\gamma\eta} = (r_{56}r_{78}/r_{57}r_{68})^{1/2}.$$

$$\begin{aligned} P_{r_{\infty\omega}r_{\infty\gamma}} = (r_{\infty\omega}r_{\infty\gamma}/4N)[&(1/r_{17}r_{34})(r_{13}r_{47}+r_{14}r_{37}) \\ &+(1/r_{12}r_{35})(r_{13}r_{25}+r_{15}r_{23})+(1/r_{13}r_{24})(r_{12}r_{34}+r_{14}r_{23}) \\ &+(1/r_{13}r_{57})(r_{15}r_{37}+r_{17}r_{35})+(1/r_{24}r_{57})(r_{25}r_{47}+r_{27}r_{45}) \\ &-(1/r_{17}r_{24})(r_{12}r_{47}+r_{14}r_{27})-(1/r_{24}r_{35})(r_{23}r_{45}+r_{25}r_{34}) \\ &-(1/r_{12}r_{57})(r_{15}r_{27}+r_{17}r_{25})-(1/r_{34}r_{57})(r_{35}r_{47}+r_{37}r_{45}) \\ &+r_{27}/r_{12}r_{17}+r_{45}/r_{34}r_{35}-r_{37}/r_{13}r_{17}-r_{15}/r_{13}r_{35} \\ &-r_{23}/r_{12}r_{13}-r_{14}/r_{13}r_{34}+1/r_{13}^2-1]. \end{aligned} \tag{28}$$

$$\begin{aligned} P_{r_{\infty\omega}r_{\gamma\eta}} = \\ (r_{\infty\omega}r_{\gamma\eta}/4N)[&(1/r_{12}r_{56})(r_{15}r_{26}+r_{16}r_{25})+(1/r_{34}r_{56})(r_{35}r_{46}+r_{36}r_{45}) \\ &+(1/r_{12}r_{78})(r_{17}r_{28}+r_{18}r_{27})+(1/r_{34}r_{78})(r_{37}r_{48}+r_{38}r_{47}) \\ &+(1/r_{13}r_{57})(r_{15}r_{37}+r_{17}r_{35})+(1/r_{24}r_{57})(r_{25}r_{47}+r_{27}r_{45}) \\ &+(1/r_{13}r_{68})(r_{16}r_{38}+r_{18}r_{36})+(1/r_{24}r_{68})(r_{26}r_{48}+r_{28}r_{46}) \\ &-(1/r_{13}r_{56})(r_{15}r_{36}+r_{16}r_{35})-(1/r_{24}r_{56})(r_{25}r_{46}+r_{26}r_{45}) \\ &-(1/r_{13}r_{78})(r_{17}r_{38}+r_{18}r_{37})-(1/r_{24}r_{78})(r_{27}r_{48}+r_{28}r_{47}) \\ &-(1/r_{12}r_{57})(r_{15}r_{27}+r_{17}r_{25})-(1/r_{34}r_{57})(r_{35}r_{47}+r_{37}r_{45}) \\ &-(1/r_{12}r_{68})(r_{16}r_{28}+r_{18}r_{26})-(1/r_{34}r_{68})(r_{36}r_{48}+r_{38}r_{46})]. \end{aligned} \tag{29}$$

If the two forms of each test are equally reliable, we may assume that $r_{12}=r_{34}$, $r_{14}=r_{23}$, $r_{17}=r_{35}$, $r_{15}=r_{37}$, $r_{16}=r_{38}$, $r_{18}=r_{36}$, $r_{47}=r_{25}$, $r_{45}=r_{27}$, $r_{46}=r_{28}$, $r_{48}=r_{26}$, $r_{56}=r_{78}$, and $r_{58}=r_{67}$. The reliability coefficients, r_{13}, r_{24}, r_{57}, and r_{68} will all be different. Designating either of the correlations of an equal pair by the subscripts of the first, we have,

$$\begin{aligned} P'_{r_{\infty\omega}r_{\infty\gamma}} = (r_{\infty\omega}r_{\infty\gamma}/4N)[&(2/r_{12}r_{17})(r_{13}r_{47}+r_{14}r_{15}) \\ &+(1/r_{13}r_{24})(r_{12}^2+r_{14}^2)+(1/r_{13}r_{57})(r_{15}^2+r_{17}^2)+(1/r_{24}r_{57})(r_{45}^2+r_{47}^2) \\ &-(2/r_{17}r_{24})(r_{12}r_{47}+r_{14}r_{45})-(2/r_{12}r_{57})(r_{15}r_{45}+r_{17}r_{47}) \\ &+2r_{45}/r_{12}r_{17}-2r_{15}/r_{13}r_{17}-2r_{14}/r_{12}r_{13}+1/r_{13}^2-1]. \end{aligned} \tag{30}$$

$$
\begin{aligned}
P'_{r_{\varpi\omega} r_{\gamma\eta}} = {} & (r_{\varpi\omega} r_{\gamma\eta}/4N)[(2/r_{12}r_{56})(r_{15}r_{48}+r_{16}r_{47}+r_{17}r_{46}+r_{18}r_{45}) \\
& +(1/r_{13}r_{57})(r_{15}^2+r_{17}^2)+(1/r_{24}r_{57})(r_{47}^2+r_{45}^2) \\
& +(1/r_{13}r_{68})(r_{16}^2+r_{18}^2)+(1/r_{24}r_{68})(r_{48}^2+r_{46}^2) \\
& -(2/r_{13}r_{56})(r_{15}r_{18}+r_{16}r_{17})-(2/r_{24}r_{56})(r_{45}r_{78}+r_{46}r_{47}) \\
& -(2/r_{12}r_{57})(r_{15}r_{45}+r_{17}r_{47})-(2/r_{12}r_{68})(r_{16}r_{46}+r_{18}r_{48})]. \qquad (31)
\end{aligned}
$$

For purposes of computation with formulas (30) and (31), the values of r_{12}, r_{14}, . . . should be taken as $(r_{12}r_{34})^{1/2}$, $(r_{14}r_{23})^{1/2}$, etc.

11.

$$
\begin{aligned}
F & = (p_{12}p_{34}p_{56}p_{78}/p_{17}p_{35}p_{46}p_{28})^{1/2} \\
& = (r_{12}r_{34}r_{56}r_{78}/r_{17}r_{35}r_{46}r_{28})^{1/2} \\
& = r_{\varpi\omega}r_{\gamma\eta}/r_{\varpi\gamma}r_{\omega\eta}. \qquad (32)
\end{aligned}
$$

Practical tetrad ratio (V, (9), (10), (19)).

$$
\begin{aligned}
\sigma_F = F(& \sigma^2_{r_{\varpi\omega}}/r^2_{\varpi\omega}+\sigma^2_{r_{\gamma\eta}}/r^2_{\gamma\eta}+\sigma^2_{r_{\varpi\gamma}}/r^2_{\varpi\gamma}+\sigma^2_{r_{\omega\eta}}/r^2_{\omega\eta} \\
& +2P_{r_{\varpi\omega} r_{\gamma\eta}}/r_{\varpi\omega}r_{\gamma\eta}+2P_{r_{\varpi\gamma} r_{\omega\eta}}/r_{\varpi\gamma}r_{\omega\eta} \\
& -2P_{r_{\varpi\omega} r_{\varpi\gamma}}/r_{\varpi\omega}r_{\varpi\gamma}-2P_{r_{\varpi\omega} r_{\varpi\eta}}/r_{\varpi\omega}r_{\varpi\eta} \\
& -2P_{r_{\varpi\gamma} r_{\gamma\eta}}/r_{\varpi\gamma}r_{\gamma\eta}-2P_{r_{\omega\eta} r_{\gamma\eta}}/r_{\omega\eta}r_{\gamma\eta})^{1/2}. \qquad (33)
\end{aligned}
$$

12.

$$
\begin{aligned}
F & = (r_{12}r_{34}r_{17}r_{35}/r_{47}r_{25})^{1/2} \\
& = (p_{12}p_{34}p_{17}p_{35}/p_{47}p_{25}\sigma_1^2\sigma_3^2)^{1/2}. \qquad (34)
\end{aligned}
$$

Proportion of the variance of either form of variable one due to the general factor, when the two forms are equally reliable. (V, (13) and (21), the latter by letting subscripts 1, i, 3, iii, 4, iv, equal 1, 3, 4, 2, 5, 7, respectively).

Since the two forms of each test are assumed to be equally reliable, $r_{12}=r_{34}$, $r_{14}=r_{23}$, $r_{17}=r_{35}$, $r_{15}=r_{37}$, $r_{47}=r_{25}$, and $r_{45}=r_{27}$. The reliability coefficients, r_{13}, r_{24}, and r_{57} will all be different. Then designating either of the correlations of an equal pair by the subscripts of the first, we have,

$$
\begin{aligned}
\sigma_F = (F/2N^{1/2})[& 2/r_{12}^2+2/r_{17}^2+2/r_{47}^2+4r_{13}^2 \\
& +(2/r_{12}^2)(r_{13}r_{24}+r_{14}^2)+(2/r_{17}^2)(r_{13}r_{57}+r_{15}^2)+(2/r_{47}^2)(r_{24}r_{57}+r_{45}^2) \\
& +(4/r_{12}r_{17})(r_{13}r_{47}+r_{14}r_{15}-r_{14}r_{45}-r_{17}r_{24}) \\
& -(4/r_{17}r_{47})(r_{12}r_{57}+r_{15}r_{45})+(8/r_{47})(r_{12}r_{15}+r_{14}r_{17}) \\
& +4(r_{45}/r_{12}r_{17}-r_{15}/r_{12}r_{47}-r_{14}/r_{17}r_{47}) \\
& -8(r_{13}r_{14}/r_{12}+r_{13}r_{15}/r_{17})-10]^{1/2}. \qquad (35)
\end{aligned}
$$

For purposes of computation, the values of r_{12}, r_{14}, . . . should be taken as $(r_{12}r_{34})^{1/2}$, $(r_{14}r_{23})^{1/2}$, etc.

13.

$F = r_{13} - F'$ (where F' is the F of (34)). (36)

Proportion of the variance of either form of variable one due to the non-chance specific factor, when the two forms are equally reliable and the theory of two factors holds (V, (15)).

Under the same assumptions as those involved in the derivation of (35), we find,

$$\sigma_F = (\sigma^2_{r_{13}} + \sigma^2_{F'} - 2P_{F'r_{13}})^{1/2}. \quad (37)$$

$$\sigma^2_{r_{13}} = (1 - r^2_{13})^2/N.$$

$\sigma^2_{F'}$ is the square of equation (35).

$$P_{F'r_{13}} = (F'r_{13}/2N)[2r_{14}/r_{12}r_{13} + 2r_{15}/r_{13}r_{17} + 2r_{12}r_{15}/r_{47} + 2r_{14}r_{17}/r_{47} - 2r_{13}r_{14}/r_{12} - 2r_{13}r_{15}/r_{17} - (2/r_{13}r_{47})(r_{12}r_{17} + r_{14}r_{15}) + 2r^2_{13} - 2]. \quad (38)$$

For purposes of computation, we must follow the same procedure as for (35).

14.

$F = 1 - r_{13}$. (39)

Proportion of the variance of either form of variable one due to the error of measurement, when the two forms are equally reliable (V, (14)).

$$\sigma_F = \sigma_{r_{13}} = (1 - r^2_{13})/N^{1/2}. \quad (40)$$

15.

$F = r_{\infty\omega} r_{\infty\gamma}/r_{\omega\gamma} = \alpha^2_1 = t_{123}$. (41)

Proportion of the "true" variance of variable one due to the general factor (V, (39), (43)). The triad.

$$\sigma_F = F(\sigma^2_{r_{\infty\omega}}/r^2_{\infty\omega} + \sigma^2_{r_{\infty\gamma}}/r^2_{\infty\gamma} + \sigma^2_{r_{\omega\gamma}}/r^2_{\omega\gamma} + 2P_{r_{\infty\omega}r_{\infty\gamma}}/r_{\infty\omega}r_{\infty\gamma} - 2P_{r_{\infty\omega}r_{\omega\gamma}}/r_{\infty\omega}r_{\omega\gamma} - 2P_{r_{\infty\gamma}r_{\omega\gamma}}/r_{\infty\gamma}r_{\omega\gamma})^{1/2}. \quad (42)$$

16.

$F = 1 - r_{\infty\omega} r_{\infty\gamma}/r_{\omega\gamma}$. (43)

Proportion of the "true" variance of variable one due to the specific factor, when the theory of two factors holds (V, (40)).

σ_F in this case is given also by (42).

The various formulas of Chapter II which deal with the sums and averages of unequally reliable tests are not included here, nor are those of Chapter V which deal with the analysis of variances, covariances, and "true" correlations for the single group-factor theory. While these formulas are of undoubted importance, they are not of such general usefulness as the ones given here, and their standard errors are of such extraordinary complexity that they would seldom, if ever, be used in any ordinary investigation.

APPENDIX I

COMPUTATION OF THE INTRACLASS CORRELATION COEFFICIENT

Suppose we have given n strictly comparable tests to each of N individuals. We first make a single frequency distribution of the nN scores. From this distribution we obtain A according to the relation,

$$A = nN \sum_{1}^{nN} X^2 - \left(\sum_{1}^{nN} X\right)^2.$$

We next obtain the sum of the n scores of each individual.

$$S_j = \sum_{1}^{n} X.$$

There will be N such sums. These are now arranged in a new frequency distribution, from which we obtain B according to the relation,

$$B = N \sum_{1}^{N} (S_j)^2 - \left[\sum_{1}^{N} (S_j)\right]^2.$$

A partial check on the computations may be obtained by noting that,

$$\sum_{1}^{N} (S_j) = \sum_{1}^{nN} X.$$

Finally, we have

$$r = [(nN-1)B - (N-1)A] \div [(n-1)B + (n-1)(N-1)A].$$

The standard error of the intraclass correlation is not a satisfactory measure, as its distribution in samples is decidedly skew. By employing a transformation devised by R. A. Fisher, however, this difficulty may be largely avoided. Let,

$$z = 1/2\log(1+(n-1)r) - 1/2\log(1-r).$$

Then,

$$\sigma_z = n^{1/2}/(2(n-1)(N-2))^{1/2}.$$

A table for finding z from r is given by Fisher (1928), who also treats the theoretical aspects of the problem at some length. Having found the standard error of z, we may add and subtract twice or three times this value from the value of z. Looking up the values of r corresponding to these, we obtain some notion of the probable limits of its chance variation.

APPENDIX II

Bibliography

Brown, W., "Some experimental results in correlation." Reports of the Sixth International Congress of Psychology, Geneva, 1909.

————"Some experimental results in the correlation of mental ability." British Jr. Psy., 3, 296-322 (1910).

————"An objective study of mathematical intelligence." Biometrika, 7, 352-367 (1910).

————"The effects of 'observational errors' and other factors upon correlation coefficients in psychology." British Jr. Psy., 6, 223-235 (1913).

Crum, W. L., "Note on the reliability of a test, with special reference to the examinations set by the College Entrance Board." Amer. Math. Monthly, 30, 296-301 (1923).

Cureton, E. E., and Dunlap, J. W., "Some effects of heterogeneity on the theory of factors." Amer. Jr. Psy., 42, 608-620 (1930).

Dodd, S. C., "The theory of factors." Psy. Rev., 35, 211-234 and 261-279 (1928).*

* This paper summarizes previous literature on this subject. No further attempt is made here to list papers on the theory of factors, aside from those which deal with errors of measurement.

Dunlap, J. W., "Race differences in the organization of numerical and verbal abilities." Archives of Psych. No. 124, 1931.

Edgerton, H. A., and Toops, H. A., "A formula for finding the average intercorrelation coefficient of unranked raw scores without solving any of the individual intercorrelations." Jr. Educ. Psy., 19, 131-138 (1928).

Filon, L. N. G., and Pearson, K., "On the probable errors of frequency constants and on the influence of random selection on variation and correlation." Phil. Trans. of the Royal Society of London, 191, 229-311 (1898).

Fisher, R. A., Statistical Methods for Research Workers, Second Edition. Oliver and Boyd, 1928.

Harris, J. A., "On the calculation of intra-class and inter-class coefficients of correlation from class moments when the number of possible combinations is large." Biometrika, 9, 446-472 (1913).

Holzinger, K. J., "An analysis of the errors in mental measurement." Jr. Educ. Psy., 14, 278-288 (1923).

Jensen, M. B., "Objective differentiation between three groups in education." Genetic Psy. Mon., 3, 5 (1928).

Kelley, T. L., "A simplified method of using scaled data for purposes of testing." Sch. and Society, 4, 34-37 and 71-75 (1916).

————"The reliability of test scores." Jr. Educ. Resch., 3, 370-379 (1921).

————Statistical Method. MacMillan, 1923.

————"Note on the reliability of a test: a reply to Dr. Crum's criticism." Jr. Educ. Psy., 15, 193-204 (1924).

————Crossroads in the Mind of Man. Stanford University Press, 1928.

Monroe, W. S., The Theory of Educational Measurements. Houghton Mifflin, 1923.

Pearson, K., "On the mathematical theory of errors of judgment, with special reference to the personal equation." Phil Trans. of the Royal Society of London, Series A, 198, 235-299 (1902).

Shen, E., "The standard error of certain estimated coefficients of correlation." Jr. Educ. Psy., 15, 462-465 (1924).

Spearman, C., "The proof and measurement of association between two things." Amer. Jr. Psy., 15, 72-101 (1904).

———" 'General intelligence,' objectively determined and measured." Amer. Jr. Psy., 15, 201-293 (1904).

———"Demonstration of formulae for true measurement of correlation." Amer. Jr. Psy., 18, 161-169 (1907).

———"Correlation calculated from faulty data." British Jr. Psy., 3, 271-295 (1910).

———The Abilities of Man. MacMillan, 1927.

Symonds, P. M., "Factors influencing test reliability." Jr. Educ. Psy., 19, 73-87 (1928).

Thorndike, E. L., "Empirical studies in the theory of measurement." Arch. Psy., No. 3 (1907)

Tryon, R. C., "The interpretation of the correlation coefficient." Psy. Rev., 36, 419-445 (1929).

———"The reliability coefficient as a per cent, with applications to correlation between abilities." Psy. Rev., 37, 140-157 (1930).

Yule, G. U., An Introduction to the Theory of Statistics, Eighth Edition. Charles Griffin, 1927.

Walker, H. M., Studies in the History of Statistical Method. Williams and Wilkins, 1929.

Wilton, J. R., "The effects of observational errors and other factors upon correlation coefficients in psychology." Jr. Exp. Ped., 2, 301-304 (1914).

Wishart, J., "The generalized product moment distribution in samples from a normal multivariate population." Biometrika, 20, 33-52 (1928).

Woodyard, E., "The effects of time upon variability." Teachers College Contributions to Education No. 216 (1926).

150.8
A673
no. 125
1931
Archives of psychology

ARCHIVES OF PSYCHOLOGY

List of numbers, continued from inside front cover

63. Race Differences in Inhibition: ALBERT L. CRANE. $1.50.
64. Individual Differences in Incidental Memory: SADIE MYERS SHELLOW. $1.25.
65. Character Traits as Factors in Intelligence Test Performance: WILLIAM M. BROWN. $1.25.
66. A Study of the Sexual Interest of Young Women: F. I. DAVENPORT. $1.25.
67. The Psychology of Confidence: WM. CLARK TROW. $1.25.
68. Experimental Studies of College Teaching: HAROLD E. JONES. $1.25.
69. The Influence of Treatment for Intestinal Toxemia on Mental and Motor Efficiency: ALICE E. PAULSEN. $1.00.
70. A Study of Suggestibility of Children: MARGARET OTIS. $1.50.
71. The Value of Praise and Reproof as Incentives for Children: ELIZABETH B. HURLOCK. $1.00.
72. Attention and Interest in Advertising: HOWARD K. NIXON. $1.25.
73. An Experimental Study of Thinking: EDNA HEIDBREDER. $1.75.
74. Estimation of Time: ROBERT AXEL. $1.00.
76. Measurement of Emotional Reactions: DAVID WECHSLER. $1.75.
77. Tested Mentality as Related to Success in Skilled Trade Training: THEODORA M. ABEL. $1.25.
78. Aggressive Behavior in a Small Social Group: E. M. RIDDLE. $1.75.
79. The Memory Value of Advertisements: EDITH R. BRANDT. $1.25.
80. A Critical Examination of Test-Scoring Methods: ROSE G. ANDERSON. $1.00.
81. Thermal Discrimination and Weber's Law: ELMER A. K. CULLER. $1.75.
82. A Correlational Analysis of Typing Proficiency: LUTON ACKERSON. $1.50.
83. Recall as a Function of Perceived Relations: CORA B. KEY. $1.25.
84. A Study of the Consistency of Rate of Work: CONSTANCE E. DOWD. $1.00.
85. An Experimental Investigation of Recovery from Work: S. L. CRAWLEY. $1.25.
86. Facilitation and Inhibition: THOMAS N. JENKINS. $1.00.
87. Variability of Performance in the Curve of Work: JAMES D. WEINLAND. $1.00.
88. A Mental Hygiene Inventory: S. DANIEL HOUSE. $1.50.
89. Mental Set and Shift: ARTHUR T. JERSILD. $1.25.
90. An Experimental Investigation of Rest Pauses: CHARLES W. MANZER. $1.25.
91. Routine and Varying Practice as Preparation for Adjustment to a New Situation: LELAND W. CRAFTS. $1.00.
92. The Development of a Standardized Animal Maze: L. H. WARNER and C. J. WARDEN. $1.25.
93. An Experimental Study of Speed and Other Factors in "Racial" Differences: OTTO KLINEBERG. $1.50.
94. The Relation of Reaction Time to Measures of Intelligence, Memory, and Learning: VERNON W. LEMMON. 80 cents.
95. Is the Latent Time in the Achilles Tendon Reflex a Criterion of Speed in Mental Reactions? GEORGE H. ROUNDS. $1.25.
96. The Predictive Value of Certain Tests of Emotional Stability as Applied to College Freshmen: EDWIN G. FLEMMING. $1.00.
97. A Vocabulary Information Test: ANGELINA L. WEEKS. $1.00.
98. The Effect of Various Temporal Arrangements of Practice on the Mastery of an Animal Maze of Moderate Complexity: SIDNEY A. COOK. 80 cents.
99. Recognition Time as a Measure of Confidence: GEORGENE HOFFMAN SEWARD. $1.00.
100. Precision and Accuracy: GEORGE W. HARTMAN. 80 cents.
101. A Group Test of Home Environment: EDITH MARIE BURDICK. $1.50.
102. The Effect of Changed Material on Ability to Do Formal Syllogistic Reasoning: MINNA CHEVES WILKINS. $1.25.
103. The Effect of Incentives on Accuracy of Discrimination Measured on the Galton Bar: HUGHBERT C. HAMILTON. $1.25.
104. An Investigation into the Validity of Norms with Special Reference to Urban and Rural Groups: MYRA E. SHIMBERG. $1.25.
105. Blood Pressure Changes in Deception: MATTHEW N. CHAPPELL. 80 cents.
106. An Experimental Comparison of Psychophysical Methods: W. N. KELLOGG. $1.25.
107. The Measurement of Verbal and Numerical Abilities: MATTHEW M. R. SCHNECK. $1.00.
108. A Perseverative Tendency in Pre-School Children. A Study in Personality Differences: HAZEL MORTON CUSHING. $1.00.
109. A Preliminary Study of the Effect of Training in Junior High School Shop Courses: L. DEWEY ANDERSON. 80 cents.
110. Music Appreciation: An Experimental Approach to its Measurement: MORTIMER JEROME ADLER. $1.50.
111. Motivation in Fashion: ELIZABETH B. HURLOCK. $1.00.
112. An Experimental Evaluation of Equality Judgments in Psychophysics: W. N. KELLOGG. $1.00.
113. Illusions in the Perception of Short Time Intervals: NATHAN ISRAELI. $.80.
114. Further Studies of the Reading-Recitation Process in Learning: E. B. SKAGGS, S. GROSSMAN, LOUISE KRUEGER, WM. C. F. KRUEGER. $.80.
115. Factors Affecting the Galvanic Reflex: ROLAND C. DAVIS. $1.00.
116. A Study of the Infant's Feeding Reactions During the First Six Months of Life: ROWENA RIPIN. 80 cents.
117. An Experiment in the Measurement of Mental Deterioration: HARRIET BABCOCK. $1.25.
118. An Experimental Investigation of the Phenomenon of Postural Persistence: LOWELL SINN SELLING. $1.00.
119. The American Council on Education Rating Scale: Its Reliability, Validity, and Use: FRANCIS FOSTER BRADSHAW. $1.00.
120. A Group Factor in Immediate Memory: ANNE ANASTASI. $1.00.
121. Individual Differences in the Sense of Humor and their Relation to Temperamental Differences: POLYXENIE KAMBOUROPOULOU. $1.00.
122. Suggestibility in the Normal and Hypnotic States: GRIFFITH W. WILLIAMS. $1.00.
123. An Analytical Study of the Conditioned Knee-Jerk: GEORGE RICHARD WENDT. $1.25.
124. Race Differences in the Organization of Numerical and Verbal Abilities: JACK W. DUNLAP. $1.25.
125. Errors of Measurement and Correlation: EDWARD E. CURETON. $1.25.

DIRECTORY OF
AMERICAN PSYCHOLOGICAL PERIODICALS

AMERICAN JOURNAL OF PSYCHOLOGY—Ithaca, N. Y.; Cornell University.
Subscription $6.50. 624 pages annually. Edited by M. F. Washburn, K. M. Dallenbach, Madison Bentley, and E. G. Boring. Quarterly. General and experimental psychology. Founded 1887.

JOURNAL OF GENETIC PSYCHOLOGY—Worcester, Mass.; Clark University Press.
Subscription $7.00. 600 pp. annually. Edited by Carl Murchison. Quarterly. Child behavior, animal behavior, comparative psychology. Founded 1891.

PSYCHOLOGICAL REVIEW—Princeton, N. J.; Psychological Review Company.
Subscription $5.50. 540 pages annually. Edited by Howard C. Warren. Bi-monthly. General psychology. Founded 1894.

PSYCHOLOGICAL MONOGRAPHS—Princeton, N. J.; Psychological Review Company.
Subscription $6.00 per vol. 500 pages. Edited by Raymond Dodge. Without fixed dates, each issue one or more researches. Founded 1895.

PSYCHOLOGICAL INDEX—Princeton, N. J.; Psychological Review Company.
Subscription $4.00. 300-400 pages. Edited by Walter S. Hunter. An annual bibliography of psychological literature. Founded 1895.

PSYCHOLOGICAL BULLETIN—Princeton, N. J.; Psychological Review Company.
Subscription $6.00. 720 pages annually. Edited by Edward S. Robinson. Monthly (10 numbers). Psychological literature. Founded 1904.

ARCHIVES OF PSYCHOLOGY—New York, N. Y.; Columbia University.
Subscription $6.00. 500 pages per volume. Edited by R. S. Woodworth. Without fixed dates, each number a single experimental study. Founded 1906.

JOURNAL OF ABNORMAL AND SOCIAL PSYCHOLOGY—Eno Hall, Princeton, N. J.; American Psychological Association. Subscription $5.00. 448 pages annually. Edited by Henry T. Moore. Quarterly. Abnormal and social. Founded 1906.

PSYCHOLOGICAL CLINIC—Philadelphia, Pa.; Psychological Clinic Press.
Subscription $3.00. 288 pages. Edited by Lightner Witner. Without fixed dates (9 numbers). Orthogenics, psychology, hygiene. Founded 1907.

PSYCHOANALYTIC REVIEW—Washington, D. C.; 3617 10th St., N. W.
Subscription $6.00. 500 pages annually. Edited by W. A. White and S. E. Jelliffe. Quarterly. Psychoanalysis. Founded 1913.

JOURNAL OF EXPERIMENTAL PSYCHOLOGY—Princeton, N. J.; Psychological Review Company.
Subscription $6.00. 500 pages annually. Edited by Samuel W. Fernberger. Bi-monthly. Experimental psychology. Founded 1916.

JOURNAL OF APPLIED PSYCHOLOGY—Baltimore, Md.; Williams & Wilkins Company.
Subscription $5.50. 400 pages annually. Edited by James P. Porter. Bi-monthly. Founded 1917.

JOURNAL OF COMPARATIVE PSYCHOLOGY—Baltimore, Md.; Williams & Wilkins Company. Subscription $5.00 per volume of 450 pages. Three volumes every two years. Edited by Knight Dunlap and Robert M. Yerkes. Founded 1921.

COMPARATIVE PSYCHOLOGY MONOGRAPHS—Baltimore, Md.; The Johns Hopkins Press.
Subscription $5.00. 400 pages per volume. Knight Dunlap, Managing Editor. Published without fixed dates, each number a single research. Founded 1922.

GENETIC PSYCHOLOGY MONOGRAPHS—Worcester, Mass.; Clark University Press. Subscription $7.00 per volume of 500-600 pages. Two volumes per year. Edited by Carl Murchison. Monthly. Each number one complete research. Child behavior, animal behavior, and comparative psychology. Founded 1925.

PSYCHOLOGICAL ABSTRACTS—Eno Hall, Princeton, N. J.; American Psychological Association.
Subscription $6.00. 700 pages annually. Edited by Walter S. Hunter. Monthly. Abstracts of psychological literature. Founded 1927.

JOURNAL OF GENERAL PSYCHOLOGY—Worcester, Mass.; Clark University Press.
Subscription $7.00. 500-600 pages annually. Edited by Carl Murchison. Quarterly. Experimental, theoretical, clinical, historical psychology. Founded 1927.

JOURNAL OF SOCIAL PSYCHOLOGY—Worcester, Mass.; Clark University Press.
Subscription $7.00. 500-600 pages annually. Edited by John Dewey and Carl Murchison. Quarterly. Political, racial, and differential psychology. Founded 1929.